



Young Children's Changing Reactions to Counterintuitive Claims

Citation

Ronfard, Samuel. 2016. Young Children's Changing Reactions to Counterintuitive Claims. Doctoral dissertation, Harvard Graduate School of Education.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:36524830>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Young Children's Changing Reactions to Counterintuitive Claims

Samuel Ronfard

Paul L. Harris
Meredith L. Rowe
Robert A. LeVine

A Thesis Presented to the Faculty
of the Graduate School of Education of Harvard University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Education

2016

© 2016
Samuel Ronfard
All Rights Reserved

To Eglantine.

Acknowledgements

I am grateful to faculty at the Harvard Graduate School of Education for providing me with a strong foundation in developmental psychology and excellent statistical training. I am particularly thankful to Paul L. Harris, my advisor, for his constant support, for his detailed and quick feedback, and for the opportunity to develop my own research interests.

This thesis was written with the support of Paul L. Harris, Meredith L. Rowe, and Robert A. Levine. To Paul, I owe most of my knowledge of child development and my knowledge of experimental design. He also taught me to be patient and to think through results multiple times before starting new studies. To Meredith and Bob, I owe my interest in cross-cultural research and individual differences. Working with Paul and Meredith, as a teaching fellow and as a collaborator on various research projects has been a wonderful experience. I hope to be able to teach my future students and advisees as much as they have taught me.

I'm thankful for the excellent statistical training I have received from Katherine Masyn, John B. Willet and Andrew Ho. I learned everything I know about statistics from them: first as student and later as a teaching fellow for their classes. They taught me more than statistical methods. They also taught me to describe results clearly and instilled in me the importance of careful analysis. I'll always remember their most important lesson: what design screws up cannot be fixed with statistics! I'm especially thankful to Andrew who trusted me to teach his class, S052, while he was on leave. His confidence in my statistical and teaching abilities gave me confidence in myself.

In addition to faculty, many fellow students provided ongoing support and contributed to my education at HGSE. Courtney Pollack and Darrick Yee who co-taught S052 with me have become close friends. I could not have taught that class without their support. I am thankful to have such great friends and colleagues. From the start to the end of my doctorate, I have also drawn strength and wisdom from my cohort mates. It's been a wonderful adventure!

Of course, I would be remiss not to acknowledge the important role my family has played. I still remember attending my father's dissertation defense as a young boy. It has no doubt influenced my decision to pursue my own research. Thank you for being a role model and for always encouraging me to pursue my own dreams. My mother has been my greatest supporter throughout my life. Your confidence helped me forward as did the strong work ethic and determination you instilled in me. My brother and sister not only provided encouragements but also helped me relax. I would also like to thank Eglantine. There is so much to thank you for. *Merci, mon amour.*

This research would not have been possible without the support of Harvard University and the Hong Kong University of Science and Technology who funded my research in Shenzhen, China. Eva E. Chen deserves a special note of recognition for helping me obtain this funding, helping me acclimate to Hong Kong, and helping me recruit and train Sze Long Poon and Kong Yan Tong whose help with data collection was invaluable. I also owe a debt of gratitude to Serena Fan without whom recruiting schools in Shenzhen would have been a much more arduous task.

Finally, thank you to the schools, parents, and children for their participation.

Table of Contents

Acknowledgements	iii
Table of contents	v
Abstract	vi
Introduction	1
Study 1	6
Study 2	36
Study 3	52
Conclusion	79
References	97
Appendix A	107
Appendix B	110
Appendix C	113
Appendix D	114
Appendix E	117
VITA	123

Abstract

This dissertation examines young children's acceptance of, memory for, and doubts about counterintuitive claims. In Study 1, children aged 3- to 5-years in the United States and China were asked to categorize hybrids whose perceptual features originated from two different animals or two different objects (75% from one and 25% from the other). At first, most children categorized the hybrids in terms of their predominant perceptual features. However, after hearing counter-perceptual categorizations by an adult, children categorized fewer hybrids in terms of their predominant features. When retested 1-to 2-weeks later, the adult's earlier counter-intuitive categorizations still impacted children's categorizations but less strongly.

In Study 2, American children aged 3- to 6-years were presented with three different-sized Russian dolls and asked to say which doll was the heaviest. Most children pointed to the biggest doll. They were then told that the smallest doll was the heaviest and that the biggest was the lightest, a claim that was false. Most children subsequently endorsed this claim. Nevertheless, when the experimenter left the room, older children were likely to check it by lifting the biggest and smallest dolls. Younger children rarely conducted such checks.

In Study 3, Chinese preschool and elementary school children were presented with five different-sized Russian dolls and asked to indicate the heaviest doll. Half of the children then heard a false, counter-intuitive claim (i.e., smallest = heaviest). The remaining children heard a claim confirming their intuitions (i.e., biggest = heaviest). Again, most children endorsed the experimenter's claim even when it was counter-

intuitive. During the experimenter's subsequent absence, elementary school children explored the dolls more if they had received counter-intuitive rather than confirming testimony. Preschool children rarely explored no matter what testimony they had received. On the experimenter's return, children who had explored the dolls were likely to reject her counter-intuitive claim.

Thus, counterintuitive claims can overturn children's beliefs but their influence fades over time and is moderated by children's opportunities to search for empirical evidence. Across two cultures, older children were more inclined than younger children to use opportunities to seek empirical evidence to check counterintuitive claims.

Introduction

In a very short amount of time, young children develop a complex understanding of the physical and social world. In this task, they are aided by their ability to learn from multiple sources of information. They can learn from their own first-hand experiences through observation, exploration, and experimentation. They can also learn by listening to what other people tell them and by asking questions. Critically, these different sources of information are sometimes consistent, but they sometimes conflict with one another; conflict is particularly likely when children are learning about hidden or unobservable phenomena. For example, to understand heliocentrism, children must override their first-hand perceptions (every day, the sun appears to move relative to a fixed earth). An important question for cognitive science and developmental psychology is how children resolve such conflicts and how children's resolution of such conflicts influences their beliefs (Gelman, 2009). Although developmental psychology has made great strides in understanding the strategies children use to identify reliable informants and reliable testimony (Harris, 2012), we know relatively little about how children integrate testimony that conflicts with their perceptions and intuitions into their existing beliefs. This dissertation addresses this topic by investigating: (1) the long-term impact of receiving socially-transmitted, counterintuitive information on children's beliefs; (2) children's search for evidence, especially observational evidence, following claims that contradict their intuitions.

Most research on children's acquisition of counterintuitive and counter-perceptual ideas has focused on the immediate impact of testimony on children's beliefs, and thus we know less about its influence in the longer term (for a review, see Lane & Harris,

2014). Understanding the longer term impact of counterintuitive testimony is important because it clarifies the role of testimony in children's acquisition of such beliefs. Conceivably, the impact of counterintuitive testimony on children's beliefs may be short-lived. A single exposure to a counterintuitive claim from a single informant may be enough to shift children's beliefs in that moment, but not days or weeks later. This is because in the long-term children's beliefs might revert to their initial judgments especially if they are presented once again with the original perceptual evidence. This would suggest that children's acquisition of counterintuitive and counter-perceptual beliefs is protracted and requires repeated exposures to counter-testimony. On the other hand, one exposure to testimony from a single informant may be enough to change children's beliefs about certain counter-perceptual or counterintuitive ideas, especially if that testimony comes from a trusted source and has some empirical support. In Study 1, I examine whether such testimony does indeed influence children's beliefs after a delay of 2- to 3-weeks.

In Study 1, it was not possible for children to gather evidence to check the counter-perceptual claim that the informant had made. For example, children did not have the opportunity to ask someone else to corroborate the claim that the informant had made; nor could they make further observations of the pictured items to check on the properties they had inferred. However, outside the laboratory children may seek such additional evidence. Indeed, there is increasing evidence that infants and preschoolers seek further evidence when they *observe* events that are counter-intuitive (Legare, 2014; Legare, Wellman, & Gelman, 2010; Stahl & Feigenson, 2015; Schulz, 2012). Thus, it is plausible that children will also seek further evidence following a *claim* that is counter-intuitive.

Whether children seek such evidence may play an important role in children's longer-term acceptance or rejection of what they were told. Alternatively, it is possible that counterintuitive claims may not trigger children's search for evidence and that children may simply either accept or reject such claims. Indeed, experimental research and mathematical models of learning suggest that instruction restricts children's exploration by reducing the number of hypotheses children consider and therefore reduces the need to collect information testing these alternative hypotheses (e.g., Bonawitz et al., 2011; Shafto, Goodman, & Frank, 2012).

Children's memory for, and evaluation of, socially-transmitted, counterintuitive information may be influenced by children's socio-cultural context in important ways. Indeed, a large body of Anthropological and Psychological research demonstrates the important role of cultural beliefs and practices in shaping parents' child rearing goals and philosophies, and, in turn, how children learn (Lancy, Gaskins, Bock, 2009; Henrich, Heine, & Norenzayan, 2010; Nielsen & Haun, 2016; Rogoff, 2003). Comparing children growing up in different cultural contexts reveals universals as well as variations in developmental outcomes and thus provides insights about the important role of context on children's development. To explore whether, and if so to what extent, context influences how children acquire socially-transmitted, counterintuitive information, this dissertation compares the longer term influence of counter-perceptual testimony on American (Study 1A) and Chinese children's belief (Study 1B) as well as the influence of counter-intuitive testimony on American (Study 2) and Chinese (Study 3) children's search for confirmatory empirical evidence following such claims. American and Chinese children were chosen as comparative samples because previous work has identified

important differences between the child rearing beliefs and practices of Chinese and American parents (e.g., Chen & French, 2008; Harkness, Mavridis, Liu, Super, 2015; Markus & Kitayama, 1991) and between the pedagogical beliefs and practices of Chinese and American preschool teachers (Wang, Elicker, McMullen, & Mao, 2008). In general, American parents tend to emphasize independence and self-expression while Chinese parents emphasize deference and social harmony (e.g., Markus & Kitayama, 1991). The relative emphasis on individuality in the U.S. is also apparent in the pedagogical beliefs and practices of American and Chinese preschool teachers. Wang et al. (2008) found that, on average, Chinese teachers were more likely to endorse teacher-structured instructional approaches than U.S. teachers. While U.S. teachers were more likely to endorse less formal, less structured, child-initiated pedagogy. These differences in American and Chinese children's experiences at home and at school are likely to influence both their memory for counter-intuitive claims and their search for confirmatory evidence following such claims.

The socialization of Chinese children may lead them to place greater confidence in the counter-intuitive claims offered by an experimenter while American children's own socialization may lead them to place greater confidence in their own intuitions. Indeed, previous research has found that Chinese-American and Chinese children are more likely to endorse counter-perceptual testimony provided by a consensus of multiple informants than European-American children (Corriveau & Harris, 2010; Corriveau, Kim, Song, & Harris, 2013). It is therefore possible, that Chinese children's greater consideration of the testimony they receive will also lead them to better encode and remember counter-intuitive claims than American children. It is unclear whether

American and Chinese children will differ in their propensity to seek evidence following the experimenter's counter-intuitive claim. Following the experimenter's counter-intuitive claim, Chinese children may be more likely to endorse the claim but nonetheless wonder whether they were right to do so. In contrast, American children may be more likely to reject the testimony of the experimenter but they may also wonder whether they were right to do so. In both cases, the experimenter's testimony may lead children to seek confirmatory evidence for the counter-intuitive claim to assuage their doubts about their evaluation of it. Cross-country differences in the endorsement but not in the testing of counter-intuitive claims would suggest important but rather subtle differences in how children approach such claims. It would suggest a difference in children's initial consideration of these claims but not necessarily a difference in their long-term acceptance of them – acceptance being determined by children's search for additional information regarding the counter-intuitive claim.

In the following sections, I present the result of these three empirical studies. Each study begins with its own introduction and includes a description of the method and analyses used in the study. I have included a discussion of the implication of the results within each study but conclude with summative discussion that brings together the results of the three studies and discusses implications for research and education.

Study 1: The influence of counter-perceptual testimony fades over time:

Evidence from Chinese and American children

Children learn through observation and from the testimony of other people. These two different sources of information are sometimes consistent with each other, but they sometimes conflict (Harris & Koenig, 2006; Lane & Harris, 2014). In cases of conflict, children are surprisingly willing to defer to counter-perceptual testimony. Gelman and Markman (1986; 1987) showed that when preschoolers are presented with a label for an animal that conflicts with its appearance, they rely on that label rather than on its appearance to make inductive inferences. For example, children were shown a drawing of a flamingo and a drawing of a bat and told that the flamingo's legs got cold at night whereas the legs of the bat got warm at night. Children were then presented with a picture of another creature which looked like a bat but which the experimenter introduced as a bird. Children often inferred that this new creature's legs got cold at night – a response consistent with the way it had been categorized by the experimenter but not with the animal's appearance.

Children are especially willing to endorse counter-perceptual claims when they are made by knowledgeable informants (e.g., Lane & Harris, 2015) and when those claims have some empirical support (e.g., Bernard, Harris, Terrier, & Clément, 2015; Chan & Tardif, 2013; Jaswal, 2004; Lopez-Mobilia & Woolley, 2016). Importantly, children will convey such counter-perceptual claims to someone else (Jaswal, Lima, & Small, 2009) suggesting that their endorsement of counter-perceptual testimony reflects belief-change rather than simple compliance. However, children's willingness to endorse counter-perceptual claims has limits. Children typically reject claims that directly contradict what they know or what they see (e.g., Clément, Koenig, & Harris, 2004; Koenig & Echols, 2003; Lane, Harris,

Gelman, Wellman, 2014; Ma & Ganea, 2009; Pea, 1982; Robinson, Champion, & Mitchell, 1999; Robinson, Mitchell, & Nye, 1995). Thus, children are not indifferent to the available perceptual evidence but under certain circumstances, they are willing to base their beliefs on counter-perceptual testimony.

The primary aim of these studies was to test the influence of counter-perceptual testimony on children's beliefs over time. Most research on children's acceptance of counter-perceptual claims has focused on the immediate impact of such claims on their beliefs. We know little about their influence in the longer term. Conceivably, the impact of a counter-perceptual claim on children's beliefs is short-lived. A single exposure to a counter-perceptual claim from a single informant may be enough to shift children's beliefs in that moment, but not days or weeks later. In the long-term, children's beliefs might revert to their initial judgments especially if they are presented once again with the original perceptual evidence. This would suggest that children's acquisition of counter-perceptual beliefs might be protracted and require repeated exposures to a counter-perceptual claim for those beliefs to take hold. On the other hand, one exposure to a counter-perceptual claim from a single informant may be enough to ensure children's continued acceptance.

To test whether the effect of counter-perceptual testimony is short-lived or more long lasting, we asked 3- to 5-year-old children to categorize hybrid pictures of animals and objects (Bernard et al., 2015; Corriveau et al., 2009c; Jaswal, 2004; Jaswal & Markman, 2007). These hybrids take 75% of their perceptual features from one animal or object and 25% of their perceptual features from a different animal or object (Jaswal et al., 2009). Initially, children typically choose to categorize these

hybrids according to the majority of their characteristics—an animal that appears to be 75% cat and 25% dog is typically judged to be a cat (Bernard et al., 2015; Corriveau et al., 2009c; Jaswal, 2004; Jaswal & Markman, 2007).

Immediately following their initial categorization of the hybrid, an informant gave children testimony that countered their initial judgments. Thus, children received counter-testimony on every trial whether or not children initially categorized a hybrid based on its majority characteristics. By providing counter-testimony whether children initially categorized the hybrids based on its majority characteristics or not we ensured that all children would receive the same amount of counter-testimony from the informant. To assess the immediate impact of such testimony on children's categorization of the hybrid, children were asked to make an inference about it. They could base this inference either on their initial judgment or on the subsequent counter-claim made by the informant. In Study 1, we asked children to categorize the hybrid pictures a second time, approximately 1-2 weeks later. In Study 2, we asked children to categorize the hybrids either at 1-2 weeks later (short delay) or 3-6 weeks later (long delay). By restricting our analyses to items that children had initially categorized based on the majority features of the hybrids, this procedure allowed us to assess the immediate and longer-term impact of *counter-perceptual* testimony on children's beliefs.

Although previous research has not examined the long-term impact of such testimony, we do know that children remember the characteristics of informants after a delay and use that information when deciding which informant to trust. Corriveau and Harris (2009a) introduced 3- and 4-year-old children to two

informants, one who labeled familiar objects correctly and one who did not. When presented with novel information by these two informants immediately after learning about their labeling abilities, children endorsed the information provided by the accurate informant over the information provided by the inaccurate informant. When these two informants presented new labels for new objects a week later, children continued to prefer the previously accurate informant's labels. By implication, children can remember the epistemic characteristics of informants, and memory for those characteristics might have both an immediate and a lingering effect on children's subsequent endorsement of claims made by that informant.

Indeed, young preschoolers appear to encode the information provided by an informant differently depending on his or her epistemic characteristics. Sabbagh and Shafman (2009) taught 4- and 5-year-old children new words for novel objects. They manipulated whether children learned these words from an informant who was knowledgeable or from an informant who was ignorant. Children were then asked to name the object (i.e., "Which one is the blicket?") or to recall what had been said (i.e., "Which one did I say is the blicket?"). Children who learned the label from the knowledgeable speaker correctly answered both types of questions immediately and after a five-minute delay. In contrast, children who learned the words from the ignorant speaker could initially recall what the informant had said but not what the object was called. Moreover, after the five-minute delay, they failed to answer both types of questions correctly. This suggests that informants' epistemic qualities moderate how strongly children encode and retain the semantic information they are taught.

Thus, a subsidiary aim of this study was to test whether the longer-term effect of counter-intuitive testimony on children's belief is moderated by characteristics of the informant. Recall that immediately following their initial categorization of the hybrid, an informant gave children testimony that countered their initial judgments – that informant was identified as a *teacher* for 4 items and as a *mother* for 4 items. We selected a teacher and a mother as informants because these informants are familiar and generally trusted by preschoolers. We decided to not have children's own mother or teacher present the testimony. Instead, to control for the effects of attachment and familiarity (Corriveau et al., 2009c; Corriveau & Harris, 2009b) children were introduced to two novel women (on video) who were designated as a “mother” and as a “teacher”, respectively. To highlight their distinctive identities and roles, we introduced the mother as “a mom who has a child the same age as you” and the teacher as “a teacher who teaches children the same age as you.” Counter-balancing (across children) which woman was designated as a mother versus a teacher ruled out the possibility that individual characteristics of either woman would drive differences in children's receptivity to teachers' versus mothers' claims. In addition, by using a video presentation, the women's vocal delivery, gestures, and facial expressions could be well matched.

By having two informants we could test whether the immediate and longer-term impact of counter-perceptual testimony is moderated by the qualities of the informant. Based on the results of Sabbagh and Shafman (2009) we expected that the immediate impact of testimony would be same whether children received testimony from the teacher or the mother. However, we tentatively hypothesized

that, in the longer-term, children would be more influenced by the teacher's testimony. We made this prediction for two reasons. First, preschoolers understand that those who teach are particularly knowledgeable (Ziv & Frye, 2004) and so they might expect teachers to be more knowledgeable than mothers. Second, both American and Chinese preschoolers affirm teachers' authority to establish social and moral rules (Tisak, Crane-Ross, & Tisak, 2000; Yau, & Smetana, 2003). These conceptions of teachers might impact children's long-term retention of their claims.

We also anticipated that children's trust in these two informants might vary across cultural contexts. Specifically, we anticipated that the sort of pedagogy that children are exposed to would influence their trust in teachers vs. mothers. Chinese preschool teachers endorse a more teacher-directed pedagogy than American preschool teachers (Wang, Elicker, McMullen, & Mao, 2008). Thus, Chinese preschoolers are likely to be exposed to more explicit and formal instruction than American preschoolers and this may make teachers' expertise more salient to Chinese preschoolers. Indeed, direct instruction is particularly memorable for preschoolers. Four-year-olds are more likely to acknowledge that they have just learned something, rather than having known it all along, when what they learned was directly taught to them rather than embedded in a story (Taylor, Esbensen, & Bennett, 1994). Moreover, previous research has found that Chinese-American children are more willing than European-American children to defer to counter-perceptual testimony (Corriveau & Harris, 2010; Corriveau, Kim, Song, & Harris, 2013). These findings have been interpreted in light of a more general cultural differences in deference to authority (Chen & French, 2008; Markus & Kitayama,

1991). Such deference that may be particularly marked for an informant explicitly identified as a teacher given that respect for teachers is an important Confucian virtue that has remained central to Chinese children's conception of learning (Li, 2012). Thus, we recruited Chinese preschoolers in Study 1A and American preschoolers in Study 1B.

In sum, the primary aim of this study was to test the longer-term influence of counter-perceptual testimony on children's beliefs by comparing its influence on children's categorization of hybrid objects and animals immediately after they receive counter-perceptual testimony and after a delay. Subsidiary aims were to explore whether any longer-term effect of testimony are influenced by the source of that testimony, and whether such effects are the same or different in China and the United States.

Study 1A

Participants

We recruited a total of 48 children from three preschools in Beijing, China (24 girls; $M_{\text{age}} = 4.78$ years, $\text{Range} = 3.44 - 5.98$). For analytic purposes, we divided these children into two age groups: 4-year-old children ($n = 29$, 15 girls; $M_{\text{age}} = 4.35$ years, $\text{Range} = 3.44 - 4.99$) and 5-year-old children ($n = 19$, 7 girls; $M_{\text{age}} = 5.44$ years, $\text{Range} = 5.01 - 5.98$). We obtained a relatively diverse sample. Parents reported having completed: some high school (23%), high school (10%), some college (23%), college (29%), and graduate school (15%). All children were Han Chinese. The surveys were completed by the children's mothers (70%), fathers (27%), or unspecified (3%). Three additional children were recruited but were not

included in our analyses because they did not complete the follow-up testing session.

Procedure

A native speaker of Mandarin tested children individually in a quiet room at their school. The protocol was translated from English to Mandarin by fluent, bilingual English-Mandarin speakers and back translated to confirm the translation. Two iPads were placed in front of the child, each displaying a picture of one of the two informants in front of a blue background. The informants were both women in their mid- to late-thirties and wore similar clothing. The experimenter said: “I know these two women. This woman is a teacher. She teaches children the same age as you. This woman is a mom. She has a boy/girl the same age as you.” Whether the teacher or the mother was introduced first and whether informant 1 or 2 was identified as the teacher or the mother was counterbalanced across children. Next, as a memory check, children were asked to point to the teacher and the mother. All children passed these comprehension checks. Children were then told: “We’re going to look at some animals and objects. Then one of these two women will tell you about them. After that I will ask you some questions.”

Children were shown four laminated hybrid-animal pictures and four laminated hybrid-object pictures, each 8.5 inches by 11 inches (Table 1, U.S.; Table 2, China). These hybrids, used in prior research, combine 75% of features of one animal or object with 25% of the features from a different animal or object (Bernard et al., 2015; Corriveau et al., 2009c; Jaswal, 2004; Jaswal & Markman, 2007). For example, the cat-dog is composed of 75% cat features and 25% dog features. We

conducted pilot testing with these hybrids to create two sets (detailed in Appendix A). Each set was composed of two animals and two objects that were similarly ambiguous across our target age-range.

Before testing each child, the experimenter shuffled these eight pictures to randomize the order of their presentation. Children were shown each picture one at a time and told: "I am going to show you a picture, do you think this is a [75% label] or a [25% label]?" The order of the labels was counterbalanced across children. After children had selected a label, the experimenter told children, "Let's ask the [teacher/mother]", and played the video of the informant naming the hybrid. For each trial, children received testimony from only one informant. The mother provided information about 4 hybrids and the teacher provided information about 4 hybrids. The four hybrids that each informant provided testimony about were counterbalanced across children. For half the children within each age group, the teacher provided information about the hybrids in set 1 and the mother provided information about the hybrids in set 2. The reverse was true for the other half of children in each age group. Because the pictures were presented in a random order, the claims made by a given informant were not blocked (i.e., they went back and forth between informants depending on which item was randomly selected). The informant always stated that the hybrid was the opposite of what the child believed. For example, if the child stated that cat-dog was "a dog" the informant in the video would say (holding the same picture of the cat-dog presented to the child): "This is a cat".

Following this testimony, children's receptivity to the informant's categorization was assessed by asking them to make an inference about the hybrid. For example, for the cat-dog, children were asked: "Do you think this animal barks or meows?" Thus, children were asked to make an inference, which they could base either on their initial judgment or on the subsequent counter-claim made by the informant (for the inference questions, see Appendix A).

To assess whether the testimony children received influenced their longer-term categorization of the hybrid, children were questioned again in a second test session. They were shown the same hybrid pictures approximately 1-2 weeks later (Range: 7 - 18 days, $M = 12$ days, $SD = 4$) and asked to categorize them once again. The experimenter shuffled the eight hybrid pictures before testing each child to randomize the order of their presentation. Children were shown each picture one at a time and told: "I am going to show you a picture, do you think this is a [75% label] or a [25% label]?" The order of the labels was counterbalanced across children. Children were not reminded that they had seen the pictures before but the experimenter was familiar to them from the initial test session.

Results

To assess the shorter- and longer-term impact of counter-perceptual testimony on children's beliefs, we restricted our analyses to the items that children initially categorized based on the hybrids' majority features, i.e., items for which they received testimony that was both inconsistent with their initial judgments *and* inconsistent with the majority of the entities' visible features. As expected, children were initially likely to categorize entities in terms of the majority of their visible

features: 5.5 items out of 8 items ($SD = 1.35$, $Range = 2 - 8$), a number significantly above chance (i.e., 4.00), $t(47) = 17.93$, $p < .001$. A repeated measures ANOVA on the number of items that children initially categorized based on the majority of the entities' visible features with the within-subject factor of Informant (2: Mother vs. Teacher) and the between-subjects factor of Age (2: 4- vs. 5-years-old) revealed no significant effect of Informant, $F(1, 47) = .27$, Age, $F(1, 47) = 1.06$, and no interaction between Informant and Age, $F(1, 47) = .06$. Thus, irrespective of age and no matter which informant the experimenter proceeded to consult, children were initially prone to base their categorization of the hybrids on their majority perceptual features.

To assess the effect of counter-perceptual testimony on children's subsequent categorization of the hybrids, both immediately and in the longer term, we coded children's category-based inferences immediately following the informant's testimony (e.g., "Do you think this animal barks or meows?"), and their category labels when they were re-questioned in the second test session (e.g., "Do you think this is a cat or a dog?"), restricting analysis to those items that children initially categorized based on the majority features of the hybrids. With respect to both judgments, children's replies were coded in terms of whether or not their responses matched the majority of the perceptual evidence present in the pictures of the hybrids. Children received 1 point for each response that continued to match the perceptual evidence and 0 points otherwise. For example, children received 1 point if they called the 75%/25% cat/dog hybrid a "cat", and 0 points if they called it a "dog".

To investigate the effect of counter-perceptual testimony on children's categorization of the hybrids in the nearer and longer term, we used a multilevel logistic regression model (using the `-xtlogit-` command in Stata 14) to regress whether children categorized an item based on the majority of its perceptual features (or not) on our predictors of interest – informant (Mother vs. Teacher), age (4- vs. 5-years-old) and time (immediately after testimony vs. 1-2 weeks later). A multilevel model was required to account for repeated measures: children were asked about multiple items at two time points. A logistic model was required to model the *probability* that children would categorize an item based on the majority of its perceptual features. The use of a multi-level logistic regression model also allowed us to statistically account for the fact that children varied in the number of items they contributed to the analyses (i.e., they varied in the number of item they had initially categorized according to the majority of the perceptual evidence, which ranged from 2-8).

First, we asked whether the probability that children categorized an item according to its majority perceptual features differed immediately after receiving counter-intuitive testimony as compared to when children were questioned in the second test session. We regressed whether children categorized an item based on its majority features on whether the test was administered immediately after receiving testimony versus after a delay (see coefficient for Delay, Table 1). Additionally, we controlled for whether the informant was identified as a mother versus as a teacher (see coefficient for Teacher, Table 1) and for whether children were 4-years versus 5-years of age (see coefficient for 5-year-olds, Table 1). We display the results of

this regression in Table 1, Model 1. We report results using odds-ratios. Post-hoc analyses are General Linear Hypothesis (GLH) tests. Thus, the coefficient on Time represents the ratio of the odds that a child categorized an item based on its majority features after a delay versus immediately after receiving testimony. A significant coefficient below 1.00 would indicate a lower probability of categorizing an item based on its majority feature after a delay versus immediately after receiving testimony. By contrast, a significant coefficient above 1.00 would indicate a higher probability of categorizing an item based on its majority feature after a delay versus immediately after receiving testimony. Inspection of Table 1, Model 1 confirms that children were significantly less likely to categorize an item using its majority perceptual features immediately after receiving the counter-perceptual testimony than when they were tested after a delay of 1-2 weeks. Thus, children were less swayed by the informant's testimony and more receptive to the hybrid's perceptual features at the second test session.

Given that children were less likely to categorize items based on the counter-perceptual testimony after a delay, we asked whether this waning effect of testimony was moderated by the source of the testimony. Thus, we added the interaction between the timing of children's judgments (immediate vs. delay) and whether the informant was identified as a mother or a teacher (see coefficient for Delay x Teacher, Table 1, Model 2). This interaction was statistically significant. We illustrate this interaction in Figure 1.

Table 1

Study 1. Multi-level logistic regression model comparing the probability that Chinese children will categorize an item based on the majority of its perceptual features

	Model 1			Model 2		
	Odds-Ratio	z-scores	95% CI	Odds-Ratio	z-score	95% CI
Delay (1-2 weeks)	2.76***	4.70	1.81, 4.21	4.25***	4.62	2.30, 7.87
Teacher	.82	-.91	.54, 1.25	1.34	.89	.70, 2.54
5-years-old	.86	-.40	.42, 1.78	.86	-.39	.41, 1.80
Delay x Teacher				.43*	-1.98	.19, .99
Constant	.27***	-4.57	.15, .47	.21***	-4.84	.11, .39
Random-Effects Variance		1.00			1.21	
Interclass Correlation		.23			.24	
X²		22.90***			26.54***	
Model df		3			4	
-2 Log Likelihood		597.14			593.20	

Note. $n = 528$; $groups = 48$.

* $p < .05$, *** $p < .001$.

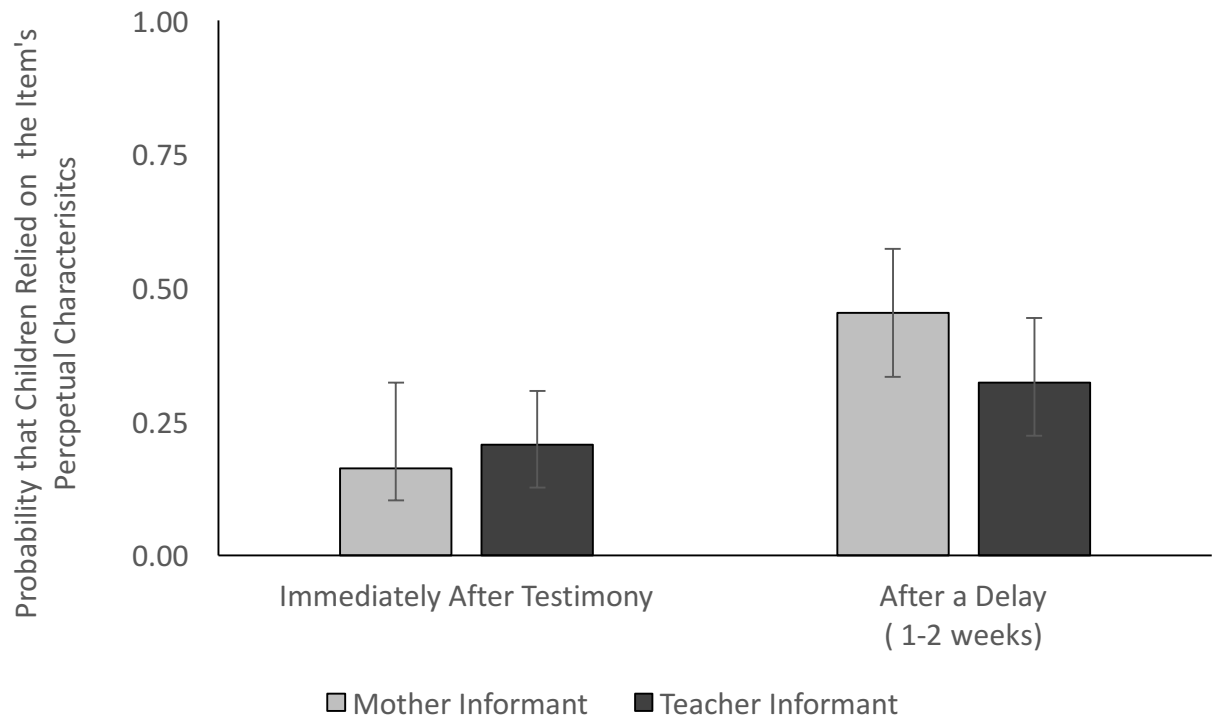


Figure 1. Probability of a Chinese child categorizing an item based on its perceptual features rather than on the counter-perceptual testimony she received from an informant identified as a mother (grey bars) or identified as a teacher (black bars) immediately after receiving testimony and after a delay of 1-2 weeks (Table 1, Model 2). Error bars represent 95% confidence intervals.

Children were equally unlikely to categorize an item according to its majority perceptual features immediately after receiving counter-perceptual testimony whether they received that testimony from the teacher or the mother. However, 1-2 weeks later, children were significantly more likely to categorize an item according to the majority of its perceptual features if they had received counter-perceptual testimony from the mother rather than from the teacher. To further explore this interaction, we compared the probability that children relied on the majority perceptual features of the item against a probability of 50%. This allowed us to test whether children were: (i) more likely to rely on the counter-perceptual testimony than on the items' perceptual features; (ii) equally likely to use these two sources of evidence; OR (iii) more likely to rely on the items' perceptual features than the counter-perceptual testimony. For items for which children received testimony from the mother, children were more likely to rely on her testimony than on the item's features immediately after receiving testimony (GLH Test: $\chi^2(1) = 30.92, p < .001$) but were equally likely to use both sources of evidence after a delay of 1-2 weeks (GLH Test: $\chi^2(1) = .54, p = .46$). For items for which children received testimony from the teacher, children were more likely to rely on her testimony than on the items' features immediately after receiving testimony (GLH Test: $\chi^2(1) = 23.36, p < .001$) and after a delay of 1-2 weeks (GLH Test: $\chi^2(1) = 8.29, p < .001$).

Thus, over time, when children received testimony from the mother they partially reverted to their initial categorization of these items based on their majority perceptual features. By contrast, when children received testimony from the teacher, they remained likely to endorse that testimony at both time points.

Finally, the effect of timing on children's judgment was not moderated by children's age (GLH Test: $\chi^2(1) = 1.21, p = .27$) and we did not find a three-way interaction between the timing of children's judgment, their age, and the identity of the informant (GLH Test: $\chi^2(3) = 5.72, p = .13$). Thus, the two-way interaction of timing and identity was not moderated by age.

Interim Discussion

Study 1A demonstrated that counter-perceptual testimony continues to influence children's belief after a delay of 1-2 weeks and that this longer-term influence is moderated by the source of that testimony. In Study 1B, we ask whether the moderating effect of informant type on children's categorization at retest is also present in the United States where preschool teachers hold more child-centered rather than teacher-directed pedagogical beliefs (Wang et al., 2008). In addition, we extend our investigation of the fading effects of counter-perceptual testimony over time by retesting approximately half of the children within roughly 1-2 weeks and retesting others within roughly 3-6 week.

Study 2

Participants

We recruited a total of 71 children from three preschools in Boston, United States (36 girls; $M_{\text{age}} = 4.81$ years, $\text{Range} = 3.48 - 5.94$). As in Study 1, we divided these children into two age groups for analytic purposes: 4-year-old children ($n = 42$, 20 girls; $M_{\text{age}} = 4.42$ years, $\text{Range} = 3.48 - 5.00$) and 5-year-old children ($n = 29$, 16 girls; $M_{\text{age}} = 5.39$ years, $\text{Range} = 5.01 - 5.94$). We obtained a relatively diverse sample. Parents reported having completed: some high school (13%), high school

(14%), some college (13%), college (22%), graduate school (34%), and did not report (4%). Children were described by their parents as White (39%), Hispanic (38%), Black (7%), Asian (3%), and other (13%). The surveys were completed by the children's mothers (87%), fathers (6%), or unspecified (8%). Eight additional children were recruited but were not included in our analyses because they did not complete the follow-up testing session.

Procedure

The procedure was identical to the one used in Study 1. Children were randomly assigned to be retested either 1-2 weeks later ($n = 33$; $M = 8$ days, Range: 4-16, $SD = 3$) or 3-6 weeks later ($n = 38$; $M = 34$ days, Range: 20-44, $SD = 8$). There was no difference in the age of children who were retested within 1-2 weeks ($M = 4.89$, $SD = .65$) and those who were retested within 3-6 weeks ($M = 4.75$, $SD = .58$), $t(69) = .97$. As with Study 1, we conducted pilot testing with the hybrids to create two equivalent sets (Appendix B).

Results

As with Study 1, we restricted our analyses to the items that children initially categorized based on the majority features of the hybrids. As expected, children were initially likely to categorize entities in terms of the majority of their visible features: 5.37 items out of 8 items ($SD = 1.47$, Range = 3 to 8), a number significantly above chance (i.e., 4.00), $t(70) = 19.34$, $p < .001$. A repeated measures ANOVA on the number of items that children initially categorized based on the majority of the entities' visible features with the within-subject factor of Informant (2: Mother vs. Teacher) and the between-subjects factor of Age (2: 4- vs. 5-years-old) revealed no

significant effect of Informant, $F(1, 77) = .99$, no effect of Age, $F(1, 77) = .05$, and no interaction between Informant and Age, $F(1, 77) = .17$. Thus, irrespective of age and no matter which informant the experimenter proceeded to consult, children were initially prone to base their categorization of the hybrids on the majority of their perceptual features.

Using the same coding scheme outlined in Study 1, we coded children's category-based inferences immediately following the informant's testimony (e.g., "Do you think this animal barks or meows?"), and their category labels when they were re-questioned in the second test session (e.g., "Do you think this is a cat or a dog?"), restricting analysis to the items that children initially categorized based on the majority features of the hybrids (which ranged from 3-8). We also used the same analytic method –multilevel logistic regression.

First, we asked whether the probability that a child categorized an item according to its majority perceptual features increased the more time passed between children's receipt of counter-perceptual testimony and their categorization of the items. That is, we tested whether children were more likely to categorize an item based on its perceptual characteristics (rather than on the testimony they received) when they were retested 3-6 weeks later rather than 1-2 weeks later, and more likely to do so when tested 1-2 weeks later rather than immediately after receiving counter-intuitive testimony. Accordingly, we regressed whether children categorized an item based on its majority features on whether the test was administered immediately after receiving testimony versus after a shorter delay (see coefficient for Shorter Delay, Table 2) and whether it was administered

immediately after receiving testimony versus after a longer delay (see coefficient for Longer Delay, Table 2). Additionally, we accounted for whether the informant was identified as a mother versus as a teacher (see coefficient for Teacher, Table 2) and for whether children were 4-years-old versus 5-years-old (see coefficient for 5-years-old, Table 2).

We display the results of this regression in Table 2. We report results using odds-ratios. Post-hoc analyses GLH tests. The first coefficient of interest is the value and statistical significance of the difference between being tested immediately after receiving counter-perceptual testimony versus being tested after a shorter delay of 1-2 weeks (see coefficients for Short Delay, Table 2). A significant coefficient below 1.00 would indicate a lower probability of categorizing an item based on its majority perceptual features after a short delay versus immediately after receiving testimony. By contrast, a significant coefficient above 1.00 would indicate a higher probability of categorizing an item based on its majority perceptual features after a delay versus immediately after receiving testimony. Other coefficients of interest are those for being tested immediately after receiving counter-perceptual testimony versus being tested after a delay of 3-6 weeks (see coefficients for Longer Delay, Table 2). Again, a significant coefficient below 1.00 would indicate a lower probability of categorizing an item based on its majority feature after a longer delay versus immediately after receiving testimony. By contrast, a significant coefficient above 1.00 would indicate a higher probability.

Inspection of Table 2 reveals that children were significantly less likely to categorize an item using its majority perceptual features immediately after

receiving the counter-perceptual testimony than when they were tested after either a short delay or a longer delay. Note that the latter coefficient is much larger than the former, indicating that children were more likely to categorize the items in terms of their perceptual features after a delay of 3-6 weeks than after a delay of 1-2 weeks. Indeed, a GLH test confirmed this interpretation, $\chi^2(1) = 11.17, p < .001$. We illustrate this pattern of results in Figure 2.

Table 2

Study 2. Multi-level logistic regression model comparing the probability that U.S. children will categorize an item based on the majority of its perceptual

	Model 1		
	Odds-Ratio	z-score	95% CI
Shorter Delay (1-2 weeks)	1.79*	2.06	1.03, 3.12
Longer Delay (3-6 weeks)	6.33***	4.75	2.96, 13.55
Teacher	.89	-.78	.65, 1.20
5-years-old	1.02	.04	.51, 2.03
Constant	.36**	-2.86	.17, 72
Random-Effects Variance		1.58	
Interclass Correlation		.32	
χ^2		23.33***	
Model df		4	
-2 Log Likelihood		1050.42	

Note. $n = 969$; $groups = 71$.

* $p < .05$, ** $p < .01$, *** $p < .001$.

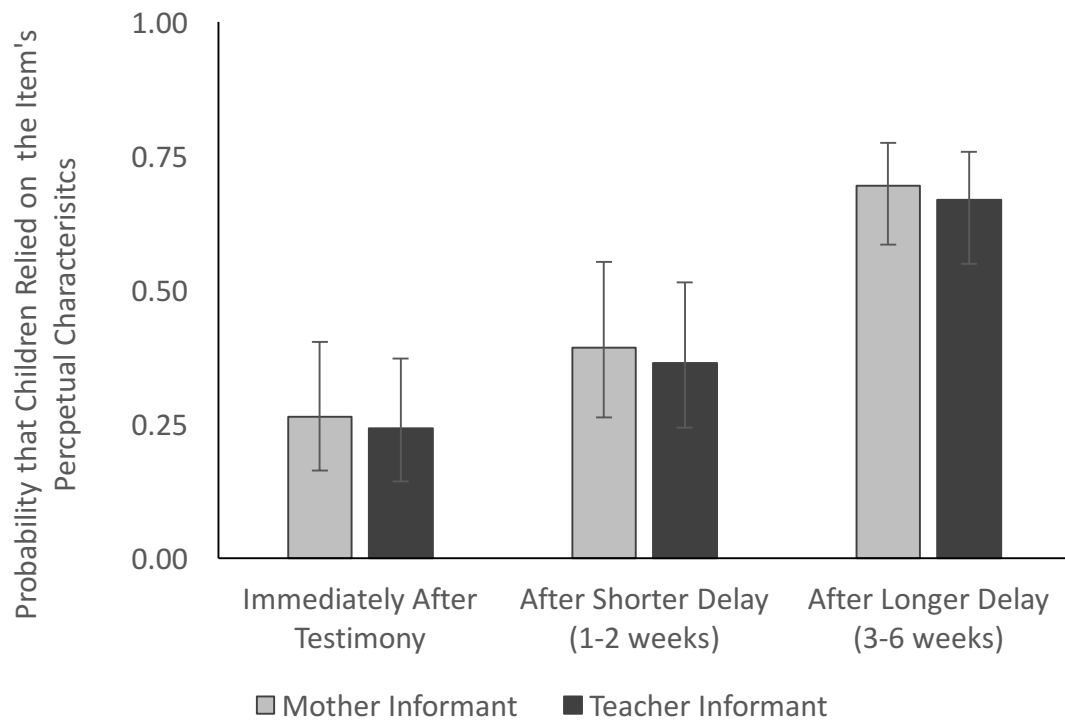


Figure 2. Probability of a U.S. child categorizing an item based on its perceptual features rather than on the counter-perceptual testimony she received from an informant identified as a mother (grey bars) or identified as a teacher (black bars) immediately after receiving testimony and after a short delay of either 1-2 weeks or a longer delay of 3-6 weeks. Error bars represent 95% confidence intervals.

In contrast to the pattern observed in Study 1, the effect of timing on children's judgment was not moderated by the informant (GLH test: $\chi^2(2) = 2.25, p = .32$). Consistent with the pattern observed in Study 1, the effect of timing on children's judgment was not moderated by children's age (GLH test: $\chi^2(2) = 2.93, p = .23$). In addition, the three-way interaction between the timing of children's judgment, their age, and the identity of the informant was not statistically significant (GLH test: $\chi^2(2) = 4.97, p = .08$).

To further explore the pattern displayed in Figure 2, we compared the probability that children relied on the majority perceptual features of the item against a probability of 50%. As in Study 1, this allowed us to test whether children were: (i) more likely to rely on the counter-perceptual testimony than on the items' perceptual features; (ii) equally likely to use these two sources of evidence; or (iii) more likely to use the perceptual features of the item than the counter-perceptual testimony. At each time point, children categorized items in a similar manner whether they received testimony from the informant identified as a mother or as a teacher. Thus, at each time point the tests were conducted on the average probability of classifying an item according to its perceptual features (collapsing across informant). When children classified items immediately after receiving counter-testimony they were more likely to rely on that testimony than on the items' features (GLH Test: $\chi^2(1) = 12.20, p < .001$). When children classified items after a short delay of 1-2 weeks, they were equally likely to use both sources of evidence (GLH Test: $\chi^2(1) = 2.85, p = .09$). Finally, when children classified items after a longer delay of 3-6 weeks, they were more likely to rely on the item's

features than on the counter-perceptual testimony they received (GLH Test: $\chi^2(1) = 10.67, p < .001$). In sum, over time, children partially reverted to their initial categorization of items based on their majority perceptual features, and the greater the amount of time between the receipt of testimony and the retest the more children reverted to their initial perceptually-based intuitions.

Discussion

Across two studies, we examined whether the effect of counter-perceptual testimony on children's beliefs is short-lived or long-lasting and whether such effects can be found in two different socio-cultural contexts. We presented children from China and the U.S. with pictures of hybrid entities whose perceptual features originated from two animals or from two objects (75% of one and 25% of the other) and asked children to categorize them. For each hybrid, children then received testimony (from an informant identified as either a teacher or a mother) that conflicted with the categories children had selected. We came back later – 1-2 weeks later in Study 1; and either 1-2 weeks or 3-6 weeks later in Study 2 – and we asked children to categorize the hybrids again. We focused our analyses on items that children initially categorized based on the majority of their perceptual features (i.e., items for which they received testimony that countered their initial beliefs and the perceptual evidence). American and Chinese children categorized fewer items according to their majority features immediately after receiving testimony. This effect persisted but it faded over time – it was somewhat weakened but still strong by 1-2 weeks, but the effect had almost completely disappeared by 3-6 weeks. Thus, with time, children increasingly reverted to categorizing the hybrids

according to their majority perceptual features. In addition, in the Chinese sample, but not in the U.S. sample, the long-term effect of the testimony was moderated by the informant. In the Chinese sample, children reverted to their initial perceptually-based categorization of the hybrid more often when they had received testimony from the mother rather than the teacher.

Children in both China and the U.S. frequently deferred to the counter-perceptual testimony they received. Thus, in both cultural settings, children were unlikely to draw inferences based on the perceptual evidence immediately following counter-perceptual testimony, a finding that is consistent with prior research (e.g., Chan & Tardif, 2013; Gelman & Markman, 1986; 1987; Jaswal et al., 2009). It is noteworthy that children in our study deferred to the informant's counter-perceptual testimony slightly more often than in Jaswal's (2004) study using the same hybrids. In that study, 4-year-olds made inferences based on the majority characteristics of 59% of the hybrids when the experimenter presented a counter-perceptual label as compared to 19% in Study 1 and 25% in Study 2. One explanation for children's greater deference in the present study is that children received testimony from informants with familiar social roles – a teacher and a mother, rather than from an informant with no designated social role. This may have increased children's willingness to defer in our study because preschoolers are likely to conceptualize both teachers and mothers as knowledgeable and trustworthy informants. Indeed, children are more likely to accept counter-intuitive claims when provided with testimony by informants who are more knowledgeable (Lane & Harris, 2015). This speculation receives support from a study by Chan and

Tardif (2013), who also used an informant identified as a “teacher” and found that U.S. preschoolers deferred about 75% of the time – a rate similar to what we found.

The counter-perceptual testimony that children received had a persistent influence. When they were retested 1-2 weeks later, children still categorized fewer hybrids based on the perceptual evidence than they had done prior to receiving counter-testimony – although they made more perception-based categorizations than they had immediately following the testimony. When children (both in China and the U.S.) were retested after a 1-2 week delay, they were equally likely to classify the hybrids based on the testimony they received and on their majority perceptual features. One could argue that such divided performance at retest means that children were guessing rather than still being influenced by the counter-perceptual testimony. However, even this conservative interpretation suggests a strong impact of counter-perceptual testimony because it implies that, even after a delay, children continued to be skeptical about the physical evidence they had initially trusted. An additional objection to our claim that the counter-perceptual testimony had a lasting influence is that the experimenter was the same for both testing sessions. Thus, it could be argued that children at the follow-up test simply responded with what they thought the experimenter wanted to hear or that children simply learned that the counter-intuitive labels should be used in the presence of the experimenter. However, this seems unlikely for two reasons. First, the experimenter never gave feedback to the children and the experimenter was not the one who provided children with the testimony. It is therefore unclear how children would infer that the experimenter preferred one label or another. Second, at least in

the Chinese sample, the long term effect of the testimony differed depending on the informant who provided the testimony. If children were making categorization decisions to please the experimenter, it is unclear why they would have selectively endorsed more counter-perceptual labels provided by the informant identified as a teacher rather than a mother.

Thus, an important implication of the two studies is that a single exposure to counter-perceptual testimony from an adult is enough to shift some of children's beliefs over time. Nevertheless, the effect of testimony did weaken in that children reverted to their initial, perceptually-based judgments the more time transpired between the receipt of testimony and retest. By implication, complete acceptance of counter-perceptual claims may require repeated exposure from a single informant or a single exposure from multiple informants to prevent children's beliefs from reverting to their initial beliefs – at least when those beliefs can be reactivated by perceptual evidence.

Admittedly, children were provided testimony that was only somewhat counter-perceptual—e.g., the cat-dog contained features of both animals and thus the counter-testimony mapped-on to at least 25% of that animal's visible features. It remains to be seen how long-lasting children's beliefs would be following dramatically counter-perceptual testimony; for example, when taught that an animal or object is completely different from what it appears to be (e.g., Lane et al., 2014).

We found that the counter-perceptual testimony from the informant identified as a teacher was longer-lasting than the testimony from the informant

identified as a mother in the Chinese sample but not in the U.S. sample. When retested, Chinese children were more likely to continue to endorse the counter-perceptual testimony if it had been provided by the informant identified as a teacher rather than by the informant identified as a mother. No such effect was found in the American sample. This pattern of results is consistent with the research reviewed in the introduction regarding differences in Chinese and American teachers' pedagogical beliefs and in cultural differences in deference to authority. However, we did not measure children's conception of the knowledge and authority of the informants. Accordingly, this explanation should be regarded as tentative; more research is needed to establish it securely. Nonetheless, in combination with the results from Sabbagh and Shafman (2009), our results suggest that future research on selective trust might benefit from the inclusion of immediate and delayed tests to assess the extent to which differences between informants become apparent over time even if they do not emerge immediately. Such delayed effects may be especially likely to emerge when children receive counter-testimony from informants who greatly differ in their expertise (see Mills, 2013). In such cases, children's perceptions and prior intuitions may act as a magnet 'pulling' children away from the testimony they have received and, in the process, revealing differences in the strength of children's encoding of that testimony based on informant characteristics.

In the present study, it was not possible for children to gather further evidence to check the counter-perceptual claim that the informant had made. For example, children did not have the opportunity to ask someone else to corroborate the claim that the informant had made; nor could they make further observations of

the pictured items to check on the properties they had inferred. However, outside of the laboratory children may seek such additional evidence. Future research might examine the extent to which children seek additional confirming or disconfirming evidence and whether such information gathering moderates the impact of a claim over the long-term.

In conclusion, we found that young children raised in two different cultural contexts display a robust and powerful ability to learn from the testimony of other people, even when that testimony conflicts with their perceptions. The impact of such testimony on children's beliefs persists a while, but slowly fades with time.

**Study 2: American children's search for evidence following claims that contradict
their intuitions**

To learn about the world, children can gather first-hand evidence based on their own observations or they can tap into the accumulated knowledge of other people by listening to what other people tell them. When learning about the distant past, about remote places, or about hidden causal processes, children cannot easily gather first-hand evidence and must typically rely on the testimony provided by other people (Harris & Koenig, 2006). Indeed, preschool children are ready to trust what they are told in these various domains (Harris, 2012; Harris & Koenig, 2006).

However, children do not believe everything they are told. They typically reject claims that directly contradict what they know or what they see (Clément, Koenig, & Harris, 2004; Jaswal, 2004; Koenig & Echols, 2003; Lane, Harris, Gelman, Wellman, 2014; Pea, 1982; Robinson, Champion, & Mitchell, 1999; Robinson, Mitchell, & Nye, 1995). For example, 16-month-old infants produce the correct labels for objects even after being given incorrect ones (Koenig & Echols, 2003); and 3- to 5-year-old children will reject an incorrect claim about an object's color even from a previously reliable informant (Clément et al., 2004).

Children are willing to endorse an informant's counter-intuitive claim if their first-hand observations leave open the possibility that the informant might be correct. For example, children are more willing to make inferences about an object or animal based on an informant's counter-intuitive claim when it has more features consistent with the informant's claim and fewer features supporting children's initial judgment (Bernard, Harris, Terrier, & Clément, 2015). Children's appreciation of the distinction between reality and appearance also influences their willingness to endorse counter-intuitive

claims. Children who have a firm grasp of this distinction are more likely to endorse an informant's unexpected claims about the true identity of the object, e.g., to accept that an object is really a bar of soap even though it looks like a rock (Lane et al., 2014). Thus, children recognize that their initial, perception-based conclusions might not be right especially when some of the empirical evidence points in a different direction or if they readily grasp how reality might not match appearance (Lane & Harris, 2014).

However, when children verbally endorse or reject a counter-intuitive claim, they may continue to have doubts about that claim. Preschoolers might endorse a counter-intuitive claim but hold on to their intuition that it is wrong. Alternatively, children might overtly reject a counter-intuitive claim but privately wonder whether it is actually correct. In either case, such uncertainty might lead children to seek additional evidence.

Indeed, infants and preschoolers seek further evidence when they observe *events* that are counter-intuitive (Legare, 2014; Legare, Wellman, & Gelman, 2010; Stahl & Feigenson, 2015; Schulz, 2012). Thus, it is plausible that children will seek further evidence following a *claim* that is counter-intuitive. Moreover, if children do gather evidence and this evidence conflicts with what they have been told, they might question their interlocutor in the hope that she can resolve the conflict (Frazier, Gelman, & Wellman, 2009). In sum, counter-intuitive testimony may set in motion a process of inquiry that culminates in children's longer-term acceptance or rejection of what they were told based on the evidence they collect and on their discussion of this evidence with their informant.

In this study, we explore the possibility that: (1) preschoolers will seek empirical evidence to check an adult's counter-intuitive claim and (2) that preschoolers who do

seek evidence will question their interlocutor about any discrepancy between the testimony they received and the empirical evidence that they gather.

We presented children ranging from 3- to 6-years-old with three different-sized Russian dolls for visual inspection. We first asked children to say which dolls they thought were the heaviest and lightest. We expected that almost all children would infer the doll's weight based on their size (i.e., biggest = heaviest) because infants as young as 9-month-old expect bigger objects to be heavier (Mounoud & Bower, 1974). Next, the experimenter told children that the smallest doll was actually the heaviest (a claim that was, in fact, false) and then re-questioned children about the dolls' relative weight. To find out if children would seek empirical evidence to test such a counter-intuitive claim, the experimenter excused himself from the room, thereby giving children an opportunity to resolve any apparent conflict between the visible size of the dolls and the experimenter's claim—by picking them up. Finally, to assess whether children's search for evidence would lead them to report back to the experimenter about the difference between his claim and the felt weight of the dolls, we also gave children an opportunity to talk with him upon his return and coded these responses. We hypothesized that older children would be more likely to endorse the counter-intuitive claim of the informant but also more likely to test it by picking up the dolls. We made this prediction because we expected that older children's greater appreciation of the difference between reality and appearance (e.g., Flavell, Flavell, & Green, 1986) would lead them to have more doubts about the apparent weight of the dolls – doubts that would increase both their acceptance of the experimenter's claims as well as their search for evidence.

Method

Participants

We recruited 70 children ranging from 3 years and 6 months to 6 years and 3 months at three Boston area preschools that also ran afterschool programs for kindergarten and first grade students (32 girls; $M_{\text{age}} = 4$ years and 11 months, $SD = 8$ month). For analytic purposes we divided children into three age groups in one year-increments: 3-year-old children ($n = 18$; 7 girls; $M_{\text{age}} = 4$ years, 0 months, *Range*: 3 years, 6 months – 4 years, 6 months), 4-year-old children ($n = 33$; 17 girls; $M_{\text{age}} = 5$ years, 0 months, *Range*: 4 years, 6 months – 5 years, 6 months), and 5-year-old children ($n = 19$; 8 girls; $M_{\text{age}} = 5$ years, 9 months, *Range*: 5 years, 6 months – 6 years, 3 months). We asked parents to report on their level of education and on their child's ethnicity. These data confirmed that we obtained a relatively diverse sample on each dimension. Parents reported having attended: high school (26%), college (36%), and graduate school (38%). Children were described by their parents as White (46%), Hispanic (30%), Black (9%), Asian (3%), and other (12%). Two additional children were tested but were not included because their initial judgments of the weight of the dolls did not match the intuition that *biggest = heaviest*.

Materials

Three Russian dolls (that had been repainted and were therefore not recognizable as Russian dolls) were used. The largest doll was green (4 x 5.5 inches), the smallest doll was white (1 x 1.75 inches), and the intermediate doll was black (3 x 4 inches).

Procedure

Children were tested in a separate room in each preschool. Children completed five phases in a fixed order: initial judgment, counter-intuitive testimony, second

judgment, opportunity to gather evidence, opportunity to talk to the experimenter. *Initial Judgment.* The three dolls were placed in front of the child. The experimenter then asked children to point to the heaviest doll and to the lightest doll: “Can you point to the one that is the heaviest?”, “Can you point to the one that is the lightest?” *Counter-Intuitive testimony.* The experimenter told children: “Actually, this one (the smallest doll) is the heaviest and this one (the largest doll) is the lightest”. *Second judgment.* Children were asked to point to the heaviest doll and to the lightest doll. In addition, they were asked to make an inference about which of the two dolls would be harder to lift: “Would it be harder for someone to pick up this object [pointing to the biggest one] or this object [pointing to the smallest one]?” *Opportunity to gather evidence.* The experimenter told children that he was going to write down some notes but that he would leave the dolls on the table. He walked outside and returned after 45s. A confederate in the room who had been reading a newspaper and had not interacted with the child (he was already reading the newspaper when children arrived in the room) recorded whether children picked up one of the dolls as well as which dolls they picked up. *Opportunity to talk to the experimenter.* Once the experimenter returned, he told children: “We’re done! Is there anything you want to tell me before we go back to the classroom?” This gave children an opportunity to tell the experimenter whether they picked up the dolls and if so, what they discovered

Results

We first examine children’s judgments about the weight of the dolls before and after the experimenter’s counter-intuitive claim. We then examine whether children sought evidence (i.e., whether they picked up the dolls or not) and, finally, whether they

reported the discrepancy between the experimenter's testimony and the felt weight of the dolls to the experimenter on his return.

Children's judgments about the weight of the dolls

Before hearing the experimenter's unexpected claim, children were asked to identify which doll was the heaviest and which one was the lightest. All children who identified the biggest doll as the heaviest and the smallest doll as the lightest (70 out of 72) were retained in the study. After being told by the experimenter that the smallest doll was actually the heaviest, children were asked three further questions about the weight of the dolls (i.e., they were asked to point to the heaviest and the lightest doll and they were also asked which of the two indicated dolls would be harder to pick up). Only a minority of children held steadfast to their initial judgment that *biggest = heaviest* for all three questions ($n = 13$, 18.57%). The remaining children either completely endorsed (3 questions out of 3; $n = 32$, 45.71%) or partially endorsed (1 or 2 questions out of 3; $n = 25$, 35.71%) the testimony of the experimenter. To evaluate the effect of the counter-intuitive testimony on children's judgments, we calculated for each child the proportion of questions for which they stated that *biggest = heaviest* before receiving testimony, $M = 1.00$ $SD = 0.00$, and after receiving testimony, $M = .35$ $SD = .38$. A Wilcoxon Signed-Ranks Test confirmed that children's judgments changed significantly following the receipt of counter-intuitive testimony, $Z = 7.12$, $p < 0.001$. To investigate whether the proportion of questions for which children stated that *biggest = heaviest* after receiving testimony was related to their age, we conducted a test for trend across ordered groups (*nptrend* command in Stata 14). We found that older children were marginally more

willing than younger children to set aside their initial intuitions and endorse the experimenter's counter-intuitive testimony, $z = 1.90$, $p = .058$ (Figure 1).

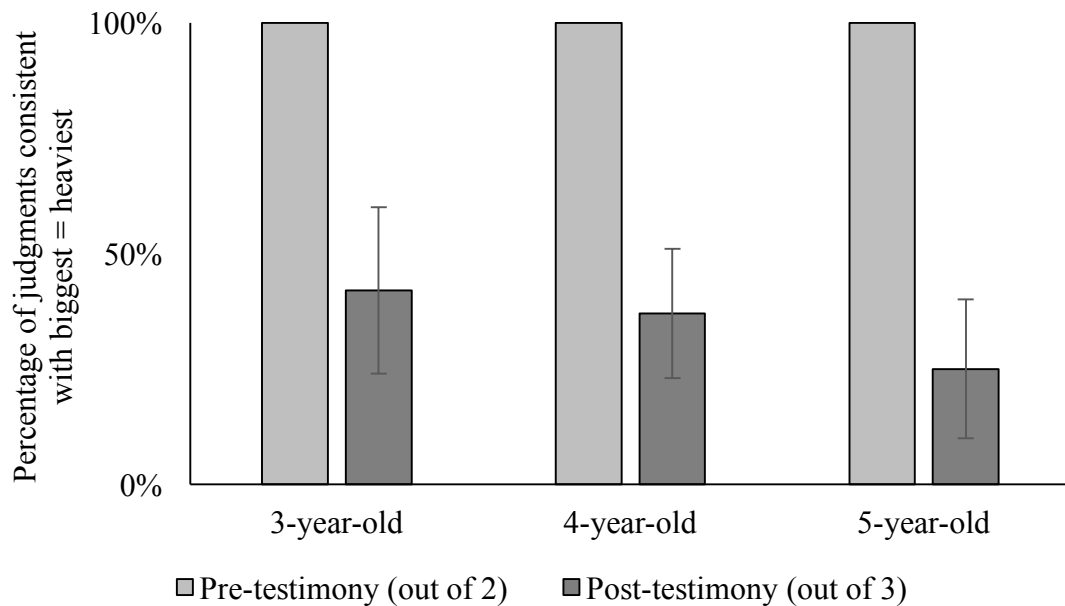


Figure 1. Percentage of judgments consistent with biggest = heaviest for 3-, 4-, and 5-year-old children before testimony and after testimony.

Children's search for evidence following testimony

After children had judged the weight of the dolls, the experimenter stated that he was going to leave to write down some notes but that he would leave the dolls on the table. The confederate then discreetly recorded whether children picked up the dolls and if so which dolls. We used this information to measure whether children sought evidence. Children were only given credit for picking up both of the two dolls that they had received testimony about (i.e., the biggest *and* the smallest doll). We considered children

to have gathered evidence if they picked up these two dolls because in so doing they could test the counterintuitive testimony they had been given. Overall, 26% (18 of 70) of children picked up both dolls. Whether children sought evidence following the receipt of counterintuitive testimony was unrelated to their prior endorsement of the experimenter's testimony: 31% (10 of 32) of children who completely endorsed the testimony, 20% (5 of 25) of children who partially endorsed the testimony, and 23% (3 of 13) of children who held steadfast to their initial judgments picked up both dolls, $\chi^2(2, N=70) = 0.99, p = 0.61$.

Inspection of Figure 2 reveals that the percentage of children seeking evidence increased with age. Specifically, more 5-year-olds sought evidence than 3-year-olds, $\chi^2(1, n=37) = 4.50, p < 0.05, V = .35$, although there was no significant difference between 4-year-old and 5-year-old children, $\chi^2(1, n=52) = 1.81, p = .18$, or between 3- and 4-year-old children, $\chi^2(1, n=51) = 1.27, p = 0.26$.

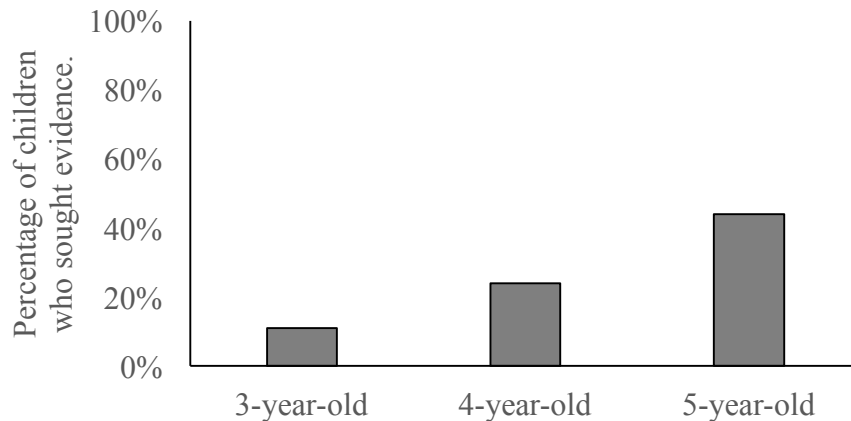


Figure 2. Percentage of 3-, 4-, and 5-year-old children who sought evidence by picking up the smallest and the biggest doll.

Children's engagement with the experimenter on his return

After the experimenter returned, he asked children whether there was anything that they wanted to tell him before returning to their classroom. Children's responses were coded by a research assistant blind to the study's goals and hypotheses and by the first author into 3 mutually exclusive categories (inter-rater agreement = 97%, disagreements were resolved through discussion). *Did not talk to the experimenter about the dolls* (71%); this category included children who either did not say anything to the experimenter or who talked about an unrelated subject (e.g., "my shirt is green and this (doll) is green"). *Sought an explanation or permission to pick up the dolls* (5%); this category included children who asked the experimenter to offer an explanation for the testimony he had provided (e.g., "Why is the white the littlest but the heaviest?") or who asked for permission or picked up the dolls after the experimenter asked whether the child wanted to say anything. *Commented on the weight of the dolls* (24%); this category included children who stated or implied that the experimenter had provided incorrect information about the weight of the largest and smallest doll (e.g., "You were tricking me, this was the heaviest one" "This one is the heaviest, I think (pointing to the largest doll)", "None of them are really heavy, they are all light").

In Figure 3, we display the percentage of children coded into each category as a function of whether or not they sought evidence. Figure 3 depicts a strong association between whether children sought evidence and tendency to comment on the weight of the dolls, Fisher's Exact Test, two sided, $p < 0.001$. Almost all of the children who did not pick up either doll failed to talk about the dolls on the experimenter's return. By contrast,

the majority of children who picked up both dolls commented on their weight upon the experimenter's return (72%).

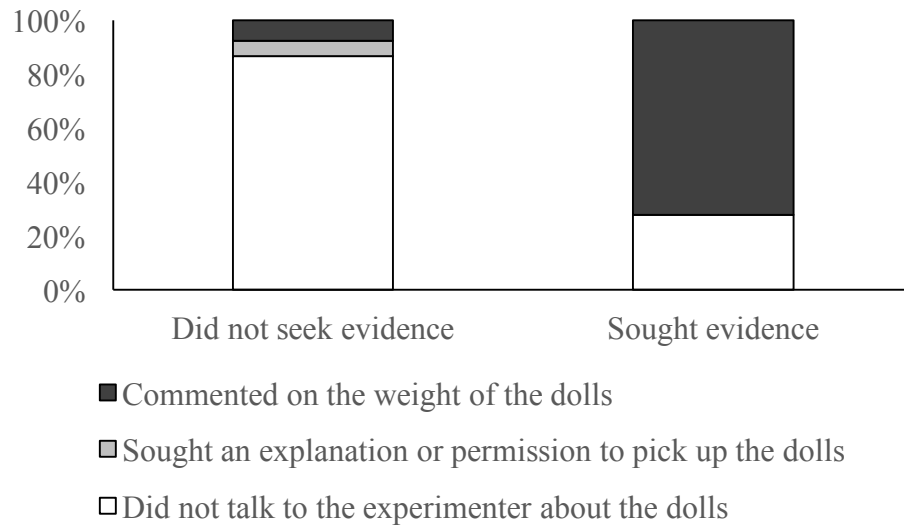


Figure 3. Percentage of children who commented on the weight of the dolls, sought an explanation or permission to pick up the dolls, or did not talk to the experimenter about the dolls upon the experimenter's return as a function of whether or not they sought evidence.

Discussion

We showed children three different-sized Russian dolls and asked them which dolls they thought were the heaviest and the lightest. We then provided them with an unexpected (and incorrect) claim about the weight of two dolls and checked whether they would endorse or reject this claim. We then gave children an opportunity to empirically test the experimenter's claim by lifting the dolls in his absence and an opportunity to discuss what they had discovered with the experimenter on his return. At the outset of

the experiment, prior to having any opportunity to lift the dolls, almost all children inferred their relative weight from their size; they judged that the biggest doll was the heaviest and that the smallest doll was the lightest. Despite this near-universal intuition, many children subsequently endorsed the contrary claim of the experimenter namely that the smallest doll was the heaviest. Indeed, older children were more prone to such endorsement than younger children. Nevertheless, the later behavior of some children indicated that, irrespective of their endorsement or denial, they were still thinking about the conflict between the appearance of the dolls and the experimenter's claim. Once the experimenter had left, a sizable minority of children – especially older children – picked up the two dolls they had received testimony about. Having done so, they often commented on the results of their investigation to the experimenter when he returned. This pattern of findings indicates that some children will seek evidence for an unexpected claim and query an informant on the basis of the evidence that they gather. We consider each of these findings in turn before discussing implications for future research.

Why were children so unanimous in their initial judgment that the biggest doll was the heaviest and yet so willing to agree with the experimenter's claim that the smallest doll was the heaviest? A plausible explanation of children's initial, unanimous, judgment is that they frequently observe a positive correlation between size and weight. Indeed, infants 9 months and older expect bigger objects to be heavier. They exert more force when lifting a bigger than a smaller object (Mounoud & Bower, 1974). However, even if children expect bigger objects to be heavier they are likely to have discovered that the correlation is imperfect, particularly as they gather more first-hand experience with

bigger but lighter objects. Indeed, older children tended to endorse the counter-intuitive claim of the experimenter more than younger children.

Following the experimenter's departure, children were left in the room with the dolls to consider two conflicting pieces of evidence: (i) the visible difference in the size and apparent weight of the dolls; and (ii) the experimenter's contrary claim. Why were older children more likely to seek evidence to resolve this conflict than younger children? Two explanations are plausible. First, older children may have experienced greater conflict than younger children between the two sources of evidence. Older children's greater appreciation that height and weight do not always co-vary might have simultaneously reduced their confidence in the perceptual evidence while increasing the plausibility of the experimenter's testimony. Younger children on the other hand might have placed greater weight on the perceptual evidence than on the testimony and thus experienced less conflict between these two sources of evidence.

A second and more speculative explanation for older children's somewhat greater acceptance of the testimony as well as their greater propensity to seek more evidence following that testimony is that they were better than younger children at generating a reason why the testimony might be true (e.g., that the smallest doll was made up of a heavier material than the biggest doll). An explanation for a counterintuitive claim makes that claim more plausible *and* also provides a way to falsify that claim if the explanation includes a testable mechanism (e.g., one of the dolls is hollow while the other has something inside it). Indeed, there is evidence that the kind of explanations children provide in response to conflicting evidence predicts their engagement in hypothesis testing behaviors (Legare, 2012) and that older children are better able to imagine

circumstances that would allow for the possibility of improbable (counterintuitive) events than younger children (Shtulman, 2009; Shtulman & Carey, 2007; Lane, Ronfard, Francioli, & Harris, 2016).

Children's subsequent willingness to report back to the experimenter about the weight of the dolls co-varied systematically with the amount of evidence they had gathered. Children almost never reported back to the experimenter if they had not picked up both dolls whereas they very often reported back to the experimenter if they did. By implication, if children collected evidence they became more confident about their initial, perception-based assessment of the dolls' relative weight and this prompted them to revisit the topic with the experimenter on his return. Children may have chosen to comment on the dolls' weight for two reasons: as a means of obtaining additional information from the experimenter about the dolls' weight or to correct the experimenter's mistaken belief about their weight. One way to discern children's motive would be to have the experimenter respond by restating his claim that the smallest doll is the heaviest. If children were seeking additional information, they would be likely to ask for an explanation from the experimenter. However, if children were correcting the experimenter, they would be likely to reaffirm their disagreement and perhaps encourage the experimenter to pick up the dolls himself.

We have argued that children's picking up of the two dolls was a search for evidence. However, it is conceivable that children picked up one or more of the dolls not to resolve the conflict between their apparent weight and what the experimenter had said, but out of idle curiosity – they simply wanted to investigate the dolls. Because we did not include a control condition in which children heard a claim that *confirmed* their intuition

about the weight of the dolls, we cannot completely rule out this possibility. However, this proposal offers no obvious explanation for the age change in children's examination of the dolls because there is no reason to expect that older children would have more curiosity about them than younger children. Moreover, the fact that children spontaneously commented on the weight of the dolls almost exclusively after they had picked them up strongly suggests that children were seeking evidence to resolve an apparent conflict rather than examining the dolls out of idle curiosity.

Our design placed a reasonable amount of social pressure on children to endorse the experimenter's testimony. Children were asked to make a second judgment about the dolls' weight immediately after receiving counter-testimony by the person who provided them with that testimony. This differs from other paradigms where children are presented with counter-perceptual or counter-intuitive testimony on video rather than in person and are not asked to make evaluative judgements by the same person who provided them with the testimony (e.g., Bernard et al., 2015; Lane et al., 2014). This additional pressure to comply may have increased children's propensity to endorse the testimony. In addition, we did not explicitly encourage children to seek out evidence when the experimenter left the room. This may have reduced children's propensity to test the experimenter's claim because children may have felt that they could not explore the dolls in his absence. Our experiment might therefore provide a conservative test of children's willingness to seek evidence following testimony.

The fact that our experimental design may have increased children's propensity to *publicly* endorse the testimony they received and reduced their exploration of the dolls might account for the lack of a relationship between children's endorsement of the

experimenter's testimony and their search for evidence. Reducing incentives for children to agree with the experimenter and cueing children to seek evidence upon the experimenter's departure from the room by using subtle or more explicit prompts might help clarify the relationship between public endorsement of testimony and children's skepticism as demonstrated by their search for evidence. Under such conditions, children who hold steadfast to their initial beliefs or completely endorse the testimony they subsequently receive might be less likely to seek evidence than children who are more uncertain.

Recent research has highlighted the tension between learning from instruction and autonomous exploration. Instruction appears to restrict exploration by reducing the number of hypotheses children consider (e.g., Shafto, Goodman, & Frank, 2012). For example, two-year-olds who received instruction engaged in less exploration of a toy and discovered fewer novel functions than children who received no instruction (Bonawitz et al., 2011; Shneidman, Gweon, Schulz, & Woodward, 2016). However, the present findings suggest that this inhibiting effect of instruction on exploration may be restricted to cases where children have no strong intuitions about the topic of instruction. In fact, our results imply that instruction can increase exploration if it conflicts with children's initial intuitions. Increased exploration following counter-intuitive claims may play an important role in children's acquisition of improbable and counter-intuitive phenomena.

**Study 3: Chinese children's search for evidence following claims that
contradict their intuitions**

Study 3 was intended to build on Study 2 and added needed controls. In Study 3, we assigned children to receive confirming or counter-intuitive testimony. This allowed us to assess whether children's exploration changed as a result of receiving counter-intuitive testimony. We also assigned children to receive a small prompt or no prompt to explore the dolls before the experimenter left the room. This allowed us to assess whether children, especially those who were younger, explored less because they did not feel comfortable doing so or because they did not think additional exploration was needed. We videotaped children's exploration of the dolls rather than having a second experimenter in the room to further reduce any inhibitions children might have felt to explore the dolls. Finally, we asked children to make additional judgements about the weight of the dolls. We asked them to make a judgment of the dolls' weight when the experimenter returned to the room. This allowed us to see whether children changed their mind about the weight of the dolls as a result of their exploration of the dolls. We also asked children to make additional judgements about the weight of the dolls with a second experimenter who was not present during the first part of the experiment. We reasoned that children would feel less pressure to continue to endorse the experimenter's testimony if they interacted with a second experimenter who was naïve about the testimony children had received about the dolls. By comparing children's responses to the questions posed by the first and the second experimenter, we could assess whether children were reluctant to contradict the first experimenter's testimony in her presence. Finally, we asked children to provide explanations for their judgements to better understand how children used the testimony and the evidence they gathered to judge the weight of the dolls.

To learn about the distant past, about remote places, or about hidden causal processes, children must typically rely on others' testimony (Harris & Koenig, 2006). A considerable body of evidence has shown that children are ready to trust what they are told in these various domains (Harris, 2012). Indeed, they often do so even when what they are told contradicts their intuitions (Lane & Harris, 2014). Children typically cannot gather empirical evidence to confirm or disconfirm what an adult has told them. For example, they cannot easily gather evidence about long departed civilizations or extinct animals. In such cases, it is reasonable for children to accept what they have been told. However, there are cases where children can test an adult's counter-intuitive claim especially when they are learning about scientific topics. For example, if presented with similarly sized cubes and told that some will float and others will sink, children can easily test the adult's claim and through their experimentation learn about the role of density. Presented with a counter-intuitive claim that is easy to test, do young children seize such opportunities or do they simply acquiesce to what they have been told?

One possibility is that young children will seek empirical evidence when presented with a counter-intuitive claim that is easily testable. After all, infants seek empirical evidence after observing counter-intuitive phenomena (Stahl & Feigenson, 2015) and preschoolers also seek evidence when faced with confounded (Schulz & Bonawitz, 2007) or theory-violating evidence (Van Schijndel, Visser, Van Bers, & Raijmakers, 2015). However, an alternative possibility is that preschoolers will not seek evidence following an adult's *claim* even if it is counter-intuitive and easily tested. Preschoolers display a robust bias to trust what other people tell them. They are willing to trust information from a previously inaccurate or misleading adult – even one who has

misled them multiple time (Jaswal, Croft, Setia, & Cole, 2010; Krogh-Jespersen & Echols, 2012; Vanderbilt, Heyman, & Liu, 2014) and even when what they are told conflicts with what they have just seen (Jaswal, 2010). This bias to trust adult testimony is especially strong in younger children (Jaswal et al., 2014). Hence, younger children may not seek evidence following a counter-intuitive claim even if older children do so.

To assess these two hypotheses, we presented preschool and elementary school children with five, different-sized, Russian dolls for visual inspection. We first asked children to say which doll was the heaviest. We expected almost all children to infer the dolls' weight based on their size (i.e., biggest = heaviest) because infants as young as 9-month-old expect bigger objects to be heavier (Mounoud & Bower, 1974). Next, the experimenter made an assertion that either confirmed their initial belief (i.e., biggest = heaviest) or contradicted it (i.e., smallest = heaviest, a claim that was, in fact, false). Children were then re-questioned about the dolls' relative weight. Next, the experimenter excused herself from the room, thereby giving children an opportunity to resolve any apparent conflict between the visible size of the dolls and the experimenter's claim – by lifting them. By comparing children's exploration of the dolls across the two conditions, we could assess the impact of counter-intuitive versus confirming testimony on children's exploration. To assess whether children's exploration would prompt them to report back to the experimenter, we also gave them an opportunity to talk with her upon her return.

Following the opportunity to explore the dolls, the experimenter again asked children about their relative weight. E1 then left the room and was replaced by a second experimenter (E2). When E2 entered the room, she expressed interest in the dolls and said she had not seen them before. She then asked children about their weight, both

directly (i.e., “Which one do you think is the heaviest?”) and indirectly (i.e., Which would make the best paperweight?”). By examining these three post-exploration judgments, we could assess the influence of children’s exploration on their subsequent judgments when they were questioned by the adult who had made the counter-intuitive or confirming claim about the dolls and also by an adult apparently uninformed about them.

Method

Participants

We recruited a total of 200 children from one preschool and from one elementary school in Shenzhen, China (101 boys; $M_{\text{age}} = 6.11$ years, $\text{Range} = 3.25 - 8.00$). The two schools serve a similar population and are located in the same neighborhood (across the street from one another). We randomly assigned preschool and elementary school children to one of four conditions: counter-intuitive testimony, counter-intuitive testimony with prompt, confirming testimony, confirming testimony with prompt (see Table 1 for descriptive statistics). An a priori power analysis informed by a pilot study determined that we needed a minimum of 37 children in each Age Group X Testimony condition to have .8 power to detect a 35% difference in children’s exploration across the two testimony conditions in each age group. We tested more children than we had originally planned because of better-than-hoped-for recruitment.

We obtained a relatively diverse sample. Parents reported on the level of education they and their partner (if applicable) had completed (196 out of 200, or 98% of parents answered this question) and on their income level (176 out of 200, or 88% of parents answered this question). Parents reported: no exposure to college (13%; i.e., neither parent had attended college), some exposure to college (24%; i.e., at least one

parent had attended college), completed college (63%; i.e., at least one parent had completed college). Parents reported having: a higher-income level (6.21%), a middle-income level (87.01%), and a lower-income level (6.21%). The surveys were completed by mothers (75.5%), fathers (22%), or were unspecified (0.50%). Thirty-one additional children were recruited but not included in our analyses because of equipment failure ($n = 20$) or because children failed to identify the largest doll as the heaviest in their initial judgment ($n = 11$).

Table 1.

Sample descriptive statistics by condition.

	Counter-Intuitive Testimony		Confirming Testimony	
	<i>Prompt</i>	<i>No Prompt</i>	<i>Prompt</i>	<i>No Prompt</i>
Preschool	$N = 18, 10 \text{ female}$	$N = 20, 9 \text{ female}$	$N = 22, 9 \text{ female}$	$N = 21, 13 \text{ female}$
	$M = 4.74, SD = 1.03$	$M = 4.49, SD = .93$	$M = 4.74, SD = .89$	$M = 4.58, SD = 1.00$
	$Range = 3.42 \text{ to } 6.25$	$Range = 3.25 \text{ to } 6.42$	$Range = 3.33 \text{ to } 6.25$	$Range = 3.33 \text{ to } 6.33$
Elementary School	$N = 31, 17 \text{ female}$	$N = 27, 12 \text{ female}$	$N = 31, 17 \text{ female}$	$N = 30, 12 \text{ female}$
	$M = 7.10, SD = .47$	$M = 7.18, SD = .43$	$M = 7.11, SD = .42$	$M = 7.07, SD = .40$
	$Range = 6.36 \text{ to } 7.87$	$Range = 6.12 \text{ to } 7.86$	$Range = 6.34 \text{ to } 8.00$	$Range = 6.09 \text{ to } 7.90$

Materials

We used five different-sized Russian dolls; each doll was attached to a square base for stability. With the square base attached, the dolls weighed: 16.32 g, 29.04 g, 46.75 g, 85.82 g, and 167.73 g. The dolls and their bases were painted white. They were arranged on a tray placed on the table so that the biggest doll was on the child's left and the smallest one was on the child's right (Figure 1). The experimenter and the child sat next to each other at the table. The dolls were approximately 18" from the table edge nearest to the child. The experimental session was discreetly recorded using a laptop camera with a darkened screen. None of the children made any comments about the laptop or behaved as if they knew they were being filmed.

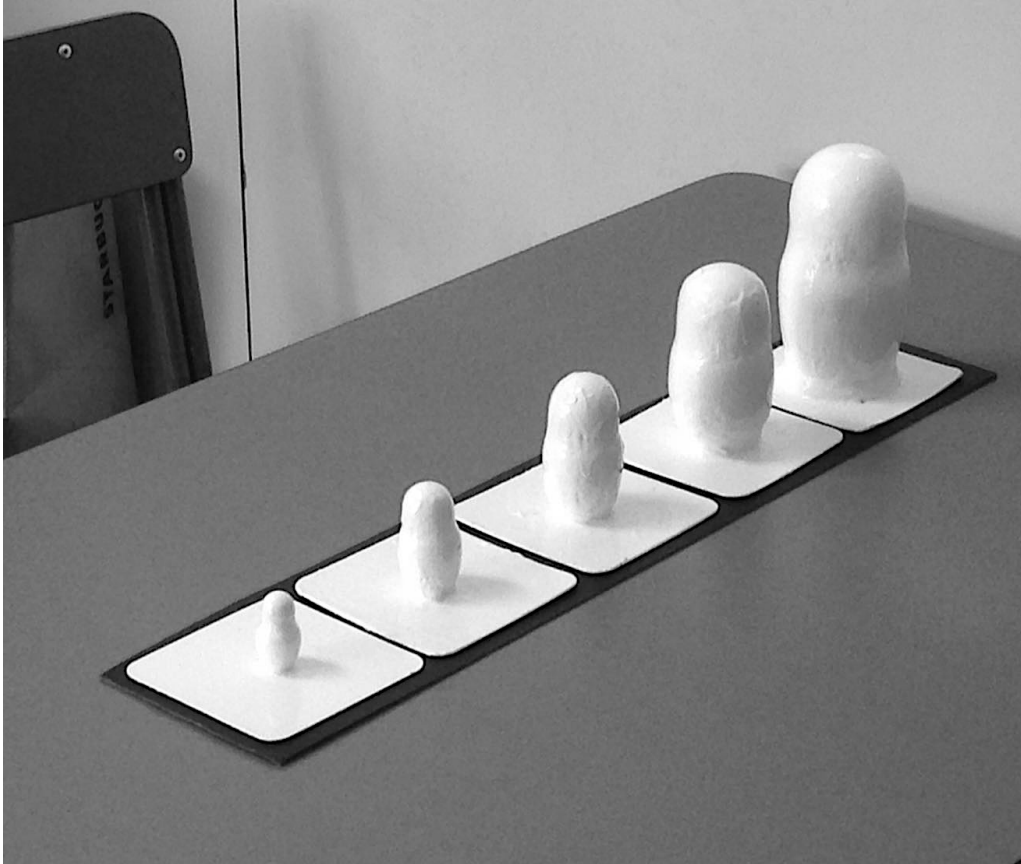


Figure 1. Stimuli used in the experiment

Procedure

Children were individually tested in a separate room at their school by a female Chinese experimenter fluent in Mandarin. The experimental procedure consisted of eight phases: (i) initial judgment with E1; (ii) testimony (counter-intuitive or confirming, depending on condition); (iii) post-testimony judgment; (iv) opportunity to explore the dolls (prompt or no prompt depending on condition); (v) opportunity to talk to the experimenter; (vi) final judgment with E1; (vii) initial judgment with E2; and (viii) paperweight selection with E2. A brief description of each phase follows.

Initial Judgment with E1. The experimenter asked children to point to the heaviest doll: “Which doll do you think is the heaviest?” Children were then asked to provide an open-ended explanation: “You think this one is the heaviest – why do you think it is the heaviest?”

Testimony. Children were randomly assigned to receive counter-intuitive testimony (i.e., smallest = heaviest) or testimony that confirmed their intuition (i.e., biggest = heaviest). In the counter-intuitive testimony condition, the experimenter told children: “Actually, that one is not the heaviest; this one here (pointing to the smallest one on the right) is the heaviest. It’s heavier than all of the other ones. It’s heavier than this one, this one, this one, this one (starting with the biggest one and moving to the second smallest one).” In the confirming testimony condition, the experimenter told children: “Yes, that one is the heaviest, and this one here (pointing to the smallest one on the right) is the lightest. This one (pointing to the biggest one) is heavier than all of the other ones. It’s heavier than this one, this one, this one, and this one (starting with the next to biggest one and moving to the smallest one)”.

Post-testimony judgment. Children were again asked to identify the heaviest doll and to provide an explanation for their judgment using the same wording as for the initial judgment. Children were also asked to recall which doll the experimenter had identified as the heaviest: “Can you point to the one I said was the heaviest?” Of the 200 children tested, 186 (93%) correctly pointed to the doll indicated by the experimenter.

Opportunity to explore the dolls. The experimenter then told children that she was going next door to pick up the phone for a moment but that she would come right back. For children assigned to the prompt condition, she added: “I’ll move the dolls a bit closer

to you” and pushed the tray so that the dolls were about 6 inches from the child. She then walked out of the room, returning after 1 minute had elapsed.

Opportunity to talk to the experimenter. Once the experimenter returned, she said, “Let’s see—we were talking about the dolls” and paused for 10 seconds to offer children an opportunity to initiate a conversation with her following their opportunity to explore the dolls. If children did not spontaneously talk to her, she prompted children: “Okay, we’re almost done. Is there anything you want to tell me?”

Final judgment with E1. Children were asked to identify the heaviest doll and to provide an explanation for their judgment using the same wording as the initial judgment. E1 then stated: “We’re almost done, can you wait in this room and somebody will come to take you back to the classroom.” She left the room and went to get E2, a different female experimenter. She did not return to the room but E2 entered the room. Note that the child had not yet met E2 who had remained in another room for the first part of the experiment.

Initial judgment with E2. E2 entered the room and exclaimed: “Whoa! I like these dolls. I’ve never played with them before! I wonder which one is the heaviest? Can you tell me?” E2 then asked for an explanation by saying: “Why do you think that one is the heaviest?”

Paperweight selection with E2. E2 then stated: “Hey—these dolls give me an idea! I think one of these dolls would be good to stop my papers from blowing away—especially if it’s heavy (E2 puts a pile of papers on the table). Can you point to the doll you think is best to stop the papers from blowing away?” After the child made a choice,

E2 asked for an explanation: “Why do you think that one is best to stop the papers from blowing away?”

Coding

The first author and a research assistant blind to the hypotheses of the study coded 25% of the total number of explanations children provided and 25% of the videos for children’s exploration of the doll. Both coders were blind to children’s age, condition, and judgments about the dolls. Agreement was 95% for the explanations and 100% for the exploration of the dolls. Disagreements were resolved through discussion. The research assistant coded the remaining explanations and videos.

Results

We analyze children’s: (i) initial and post-testimony weight judgments; (ii) exploration of the dolls; (iii) post-exploration remarks to the experimenter; (iv) post-exploration weight judgments; (v) and post-exploration weight judgments as a function of children’s exploration. We report results for logistic regression models using odds-ratios. All logistic analyses were conducted using the `–logit–` command in Stata 14.

Children’s initial and post-testimony weight judgments

In Table 2, we display the proportion of children who stated that the biggest doll was the heaviest when they were initially asked and immediately after they had heard the experimenter’s confirming or counter-intuitive testimony. All children initially stated that the biggest doll was the heaviest. Note, however, that this unanimous pattern reflects the fact that we excluded the very small minority of children ($N = 11$, i.e., 5.5%) who did not initially select the biggest doll as the heaviest. Inspection of Table 1 shows that the proportion of children who claimed that biggest = heaviest decreased sharply in both age

groups immediately after counter-intuitive testimony, McNemar tests $< .0001$, but not after confirming testimony, McNemar tests $> .25$. Children who received counter-intuitive testimony endorsed the biggest doll as the heaviest significantly less often than children who received confirming testimony, $\chi^2(1, N = 200) = 141.42$, $p < .001$, Cramér's $V = .88$. In sum, the type of testimony markedly affected children's judgments of the doll's weight.

Table 2. Percentage of children who claimed that the biggest doll was the heaviest.

		Initial Judgment	Following Testimony
Confirming	<i>Preschool</i>	100%	93%
	<i>Elementary</i>	100%	97%
Counter	<i>Preschool</i>	100%	13%
	<i>Elementary</i>	100%	10%

We coded children's explanations following their initial and post-testimony judgments. Explanations were coded as: *Bigger = Heavier* if children described a positive association between size and weight (e.g., "It's the largest so it's the heaviest"); as *Smaller = Heavier* if they described a negative association between size and weight (e.g., "It's small so it's the heaviest) or referred to being told that the smallest was the heaviest (e.g., "Because you told me"); as *Size Sometimes Unrelated to Weight* if they described why the biggest doll might not be the heaviest (e.g., "The largest one is hollow, while the smallest one is solid"). Finally, explanations were coded as *Other* if they could not be coded into the other four categories (e.g., "It's the heaviest", "It's very heavy, as heavy as four water bottles", "It is just right") or if children did not provide an

explanation. Table 3 shows the percentage of each type of explanation as a function of age, testimony type, and timing.

Table 3. Percentage of preschool and elementary school children’s explanations coded into each category as a function of Testimony Type (Confirming vs. Counter-Intuitive), and Timing (Before vs. After Testimony).

Explanation Type	Confirming Testimony		Counter-Intuitive Testimony	
	<i>Before</i>	<i>After</i>	<i>Before</i>	<i>After</i>
Preschool Children				
Other	28%	24%	24%	37%
Bigger = Heavier	79%	72%	76%	8%
Smaller = Heavier	0%	0%	0%	52%
Size Sometimes Unrelated to Weight	0%	0%	0%	3%
Elementary School Children				
Other	0%	7%	0%	10%
Bigger = Heavier	100%	82%	100%	0%
Smaller = Heavier	0%	2%	0%	50%
Size Sometimes Unrelated to Weight	0%	10%	0%	40%

Both age groups mostly offered Bigger = Heavier explanations (e.g., “It’s the largest so it’s the heaviest”) before and after confirming testimony. By contrast, both age groups mostly offered Bigger = Heavier explanations before but rarely after counter-intuitive testimony. Children who endorsed the experimenter’s counter-intuitive claim justified their decision either by repeating the experimenter’s claim (i.e., *Smaller = Heavier*) or by noting that size and weight are not always correlated (i.e., *Size is Sometimes Unrelated to Weight*, 47%). The latter explanation was common among

elementary school children (40%) but was rare among preschool children (3%). Thus, although preschool and elementary school children endorsed the counter-intuitive testimony of the experimenter at similar rates, they often did so for different reasons.

Children's exploration of the dolls

We coded whether children touched any of the dolls, and if they did, whether they gathered evidence that could confirm or disconfirm the experimenter's claim, i.e., whether or not they lifted the smallest and the biggest doll during the experimenter's absence. We further distinguished between children who picked up the smallest and the biggest doll separately and those children who picked them up concurrently, i.e., picked up the smallest doll in one hand and the biggest doll in the other.

Our analytic strategy for these two measures was to first introduce our main predictors of interest, Age Group (Elementary vs. Preschool), Type of Testimony (Counter-Intuitive vs. Confirming), Prompt (Prompt vs. no Prompt) before testing for interaction effects.

Table 3 displays parameter estimates for a logistic regression model predicting whether children touched any of the dolls. All coefficients are in odds-ratios. For example, the coefficient for Counter-Intuitive Testimony is the ratio of the odds that a child touched any of the dolls after having received counter-intuitive rather than confirming testimony.

A significant interaction between Age Group and Prompt emerged, as displayed in Figure 2. We used General Linear Hypothesis (GLH) tests to interpret this interaction. Preschoolers were as likely to touch the dolls whether they did or did not receive a prompt, GLH Test: $\chi^2(1) = .27, p > .25$. In contrast, elementary school children were

more likely to touch the dolls when they had received a prompt, GLH Test: $\chi^2(1) = 5.72$, $p = .017$. Elementary school children were more likely to touch the dolls than preschool children in the no prompt condition, GLH Test: $\chi^2(1) = 3.89$, $p = .049$, and this was especially true in the prompt condition, GLH Test: $\chi^2(1) = 17.89$, $p < .001$. There were no further interactions or main effect. Thus, older children touched the dolls more often than younger children especially when they were given a prompt to do so by the experimenter. Importantly, younger and older children's touching of the dolls was not related to the type of testimony they received. A similar proportion of children touched the dolls whether they received confirming or counter-intuitive testimony.

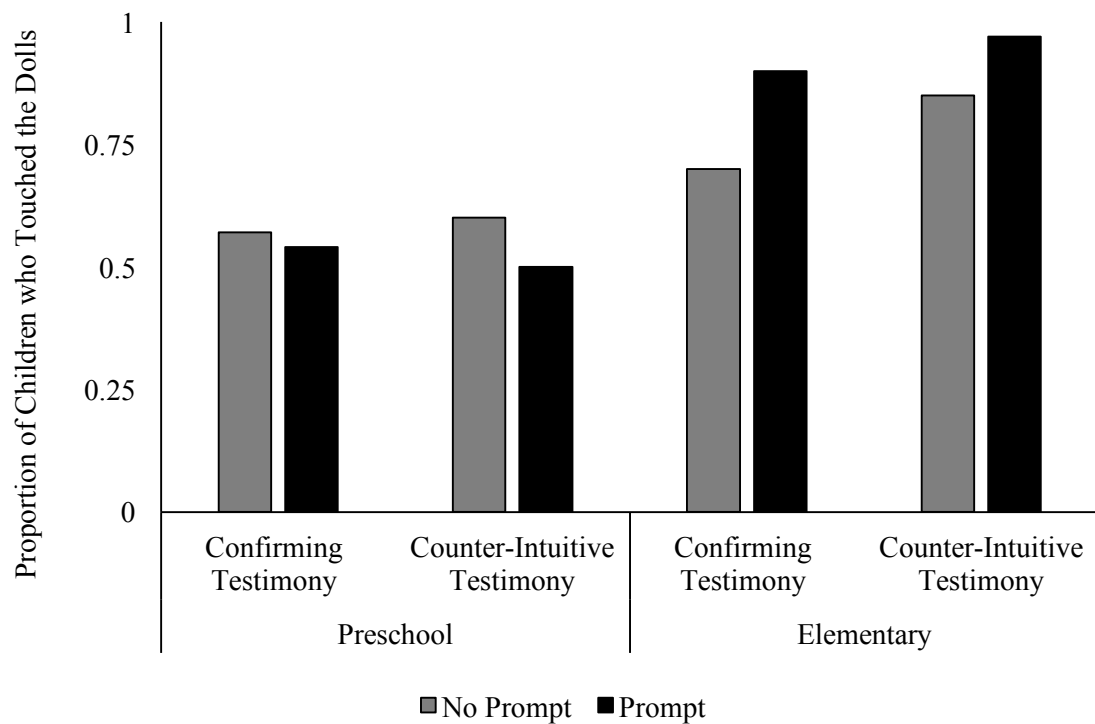


Figure 2. Proportion of preschool and elementary school children who touched the dolls while E1 was out of the room as a function of the testimony and prompting they received prior to E1's departure.

Table 3. Logistic regression model comparing whether children touched the dolls as a function of the type of testimony children received, whether they received a prompt, and their age.

	Model 2		
	Odds-Ratios	z scores	95% CI
Counter-Intuitive Testimony	1.43	1.03	.72, 2.83
Elementary	2.42*	1.97	1.00, 5.85
Prompt	.79	.52	.33, 1.91
Elementary X Prompt	5.39*	2.23	1.22, 23.66
Constant	1.19	.48	.59, 2.40
X²		30.50***	
Model df		4	
-2 Log Likelihood		200.79	

* $p < .05$; ** $p = .011$; *** $p < .001$. *Note.* $n = 200$.

We now turn to children's targeted exploration of the dolls (i.e., whether they picked up the smallest and the biggest doll). Table 4 displays parameter estimates for a logistic regression model predicting whether children picked up the biggest and the smallest doll.

A significant interaction between Age Group and Testimony Type emerged, as displayed in Figure 3. Preschoolers' exploration did not differ by type of testimony (GLH Test: $\chi^2(1) = .06, p = .80$). However, significantly more elementary school children explored following counter-intuitive than confirming testimony (GLH Test: $\chi^2(1) = 11.79, p < .001$). Thus, more elementary school children than preschool children explored following counter-intuitive testimony (GLH Test: $\chi^2(1) = 21.96, p < .001$) whereas there was no significant age difference in exploration following confirming testimony (GLH

Test: $\chi^2(1) = 2.89, p = .09$). There were no further interactions. This same pattern emerged when we focused on the subset of children who had picked up the biggest and the smallest doll at the same time (See Appendix C).

In summary, counter-intuitive testimony provoked elementary school children to compare the biggest and smallest dolls. Indeed, they engaged in such exploration whether or not they were prompted to do so by E1's placement of the dolls. This effect of counter-intuitive testimony was not observed among preschool children. Children's decision to explore was unrelated to whether they had endorsed or rejected the experimenter's testimony or to the type of explanation they had provided following their endorsement or rejection of the experimenter's claim (See Appendix D).

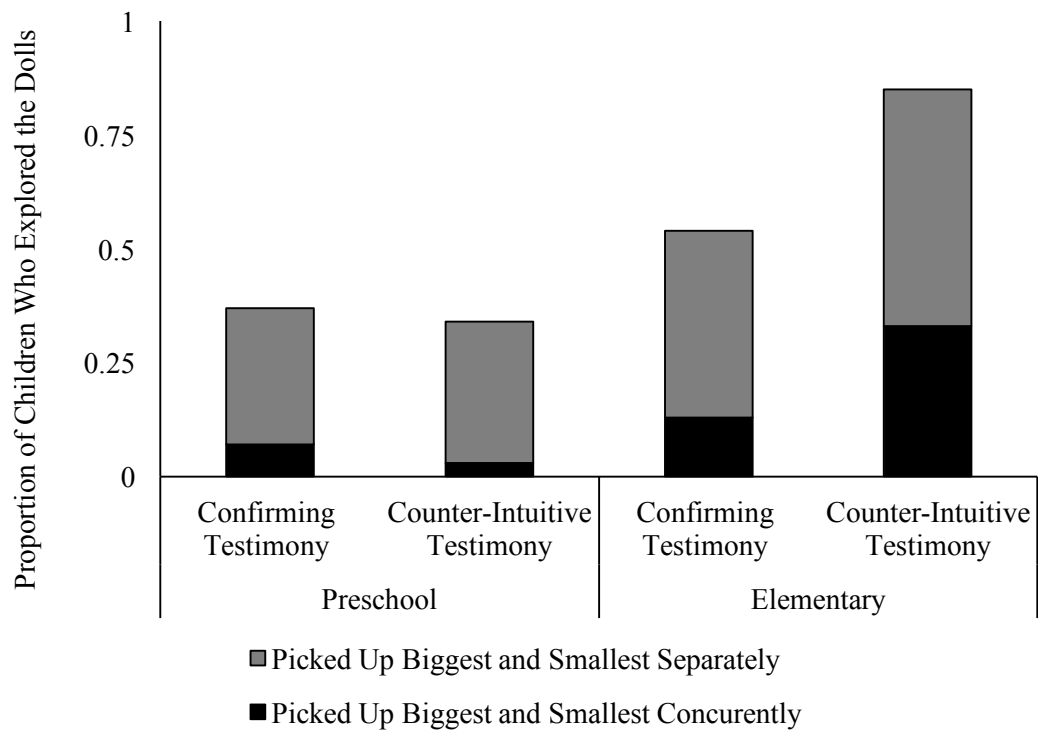


Figure 3. Proportion of children receiving confirming vs. counter-intuitive testimony who picked up the biggest and the smallest doll while E1 was out of the room.

Table 4. Logistic regression model comparing whether children picked up the biggest and smallest dolls as a function of the type of testimony children received, whether they received a prompt, and their age.

	Model 2		
	Odds-Ratios	z scores	95% CI
Counter-Intuitive Testimony	.89	.25	.36, 2.22
Elementary	2.00	1.70	.90, 4.46
Prompt	1.43	1.15	.78, 2.64
Counter-Intuitive X Elementary	5.20**	2.55	1.47, 18.43
Constant	.05*	1.99	.24, .99
X²		36.35***	
Model df		4	
-2 Log Likelihood		238.48	

* $p < .05$; ** $p = .011$; *** $p < .001$. *Note.* $n = 200$.

Post-exploration remarks to the experimenter

Following the return of E1, some children stated or implied that the biggest doll was indeed the heaviest (e.g., “The smallest doll is the lightest”; “I know which doll is the heaviest! (pointing to the largest one). I have picked up all of the dolls”). Although rare, these assertions displayed a similar pattern to children’s doll exploration.

Preschoolers seldom made them following either type of testimony (2 vs. 2, one sided Binomial Test, $p > .25$) whereas elementary school children made them more often following counter-intuitive as compared to confirming testimony (9 vs. 2, one sided Binomial Test, $p = .033$).

Post-exploration weight judgments

Children made a judgment about the weight of the dolls: (i) immediately after the opportunity to explore the dolls; (ii) when explicitly asked by E2; and (iii) when invited by E2 to select the heaviest paperweight. Preliminary analyses showed that children's judgments remained stable across these three successive time points (Appendix E). Accordingly, we summed the three judgments to reflect how often children asserted that biggest = heaviest. Children's scores were analyzed via a 2 x 2 x 2 ANOVA with Age Group (2: Preschool, Elementary), Prompting (2: Prompt, No Prompt), and Testimony Type (2: Confirming, Counter-Intuitive) as between-subject factors.

This analysis revealed significant main effects of Testimony Type, $F(1,192) = 99.46, p < .001, \eta^2 = .33$, and Age Group, $F(1,192) = 5.16, p = .024, \eta^2 = .017$. There were no other significant main effects or interaction effects. Thus, the interaction of Testimony Type X Age Group fell short of significance, $F(1,192) = 2.03, p = .16, \eta^2 = .007$. Nevertheless, we explored the simple effect of age for each type of testimony given the aforementioned age-related differences in children's exploration of dolls. In Figure 4, we display the frequency with which children said that the biggest doll was the heaviest as a function of Age Group and Testimony Type. Inspection of Figure 2 reveals that preschool and elementary school children who received confirming testimony made similar judgments, $p < .25$. In contrast, in the counter-intuitive condition, elementary school children judged that the biggest doll was the heaviest more often than younger children, $p = .011$.

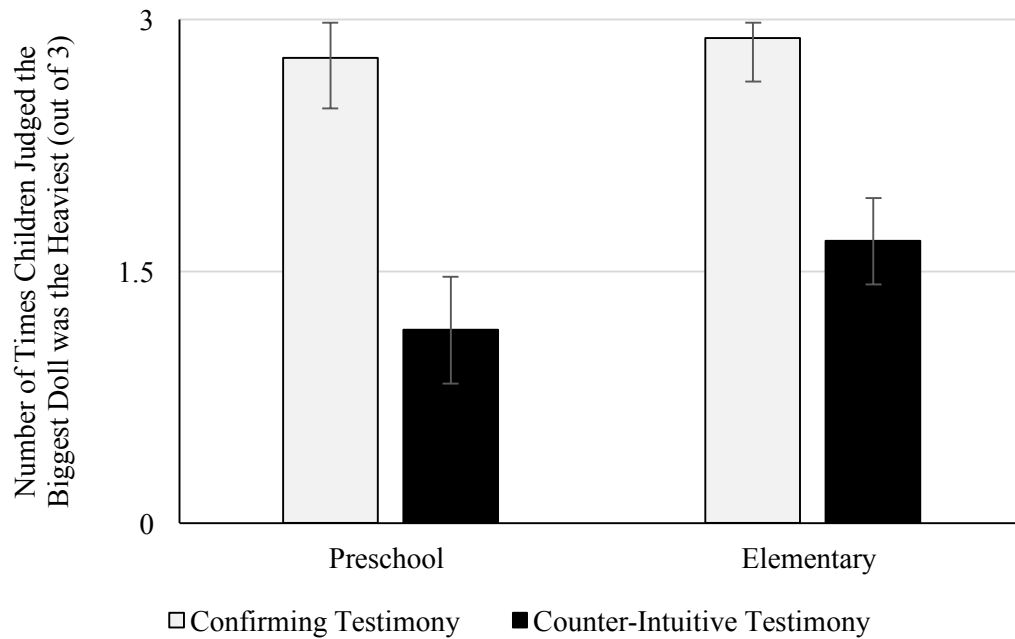


Figure 4. Average number of times preschool and elementary school children judged that the biggest doll was the heaviest as a function of the type of testimony they received. Error bars represent +/- 2 standard errors.

Post-exploration weight judgments as a function of children's exploration

To investigate the influence of exploration on children's post-exploration weight judgments, we analyzed these judgments via a 2 x 2 x 2 ANOVA with Age Group (2: Preschool, Elementary), Prompting (2: Prompt, No Prompt), and Exploration (2: Explored, Did Not Explore) as between-subject factors, restricting our analysis to children who had received counter-intuitive testimony.

This analysis revealed only a significant main effect of Exploration, $F(1,88) = 20.91, p < .001, \eta^2 = .18$: Children who had explored judged the biggest doll to be the heaviest much more often than children who had not explored. We display this main effect in Figure 5. Children who did not explore the dolls, endorsed the *smallest* doll as

the heaviest significantly above chance, $t(33) = 4.42, p < 0.001, d = 1.54$. In contrast, children who did explore the dolls, endorsed the *biggest* doll as the heaviest significantly above chance, $t(61) = 2.77, p < 0.01, d = .71$. Thus, when children gathered empirical evidence it undermined the earlier impact of the experimenter's counter-intuitive testimony on their judgments. The absence of any interaction between age and exploration, $F(1,88) = .26, p > .25$, indicates that exploration, when it occurred, shifted the judgments of preschool and elementary school children to the same extent, as shown in Figure 3.

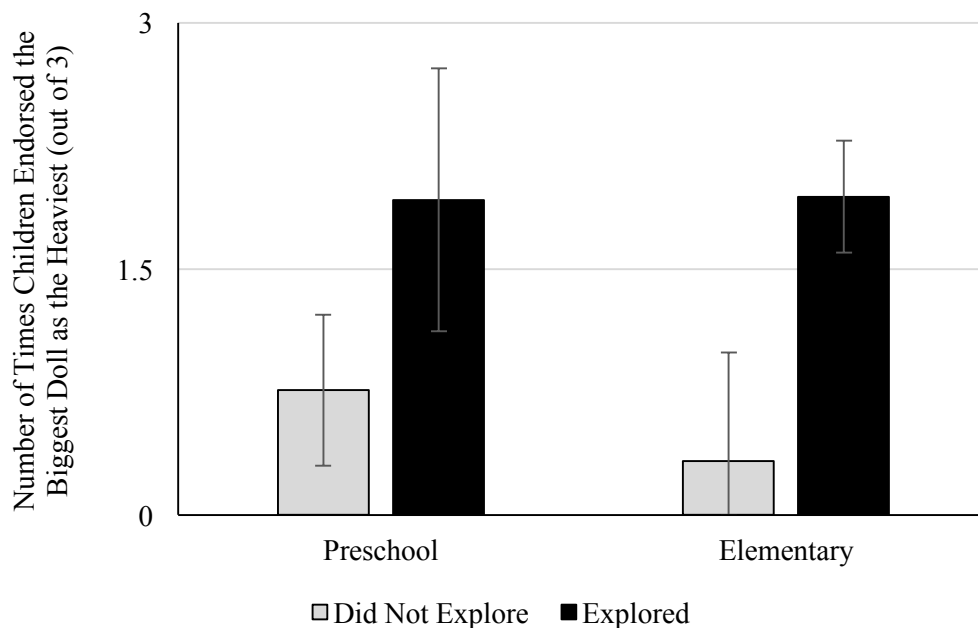


Figure 5. Average number of times preschool and elementary children judged that the biggest doll was the heaviest as a function of whether they explored the doll. Error bars represent +/- 2 standard errors.

Discussion

We asked whether children seek empirical evidence following an experimenter's testable claim. Preschool children rarely did so whether the claim matched their intuitions or conflicted with them. By contrast, elementary school children did seek empirical evidence, especially when the claim was counter-intuitive. Indeed, we found that it was children's targeted exploration of the dolls (i.e., picking up the biggest and the smallest doll) and not their more general engagement with any of the dolls that was influenced by the counter-intuitive nature of the testimony. In addition, we replicated this same age-change using a still more focused measure of children's exploration – notably, whether they picked up the biggest and the smallest doll at the same time, thereby optimizing their opportunity to establish which doll was heavier. Our favored explanation for this age change is that older children are increasingly prone to adopt an empirical stance in relation to counter-intuitive claims. Recognizing that such claims can be tested against the available evidence, they actively seek that evidence and revise their judgment of the unexpected claim accordingly. This interpretation is consistent with earlier findings showing that preschool children are surprisingly trusting of an informant's testimony, even when the informant has a history of inaccuracy and even when the claim does not match immediately prior observations. Yet it extends those findings by showing that – given relevant opportunities – older children will test unexpected claims.

Three alternative explanations for this age change received no support. First, there was no indication that the two age groups differed in their initial intuitions – almost all children in both age groups claimed that the biggest doll was the heaviest. By implication, children in both age groups found the experimenter's counter-intuitive claim equally disconcerting. Furthermore, even though there was an age change in the way that

children explained their endorsement of the counter-intuitive claim – with a considerable proportion of older children noting that size is sometimes unrelated to weight – children’s mode of explanations proved to be unrelated to their pattern of exploration. Taken together, these findings imply that older children explored more than younger children because they were better able to recognize how evidence could bear on the experimenter’s claim and not because they found the claim more unexpected or interpreted it differently.

Second, there was no indication that the two age groups varied in their ability to learn from empirical evidence, once it had been gathered. Recall that children in each age group were equally likely to maintain – in the wake of exploration – that the biggest doll was the heaviest (See Figure 5). By implication, the two age groups did not differ in their capacity to revise their judgment once empirical evidence had been gathered – rather they differed in their readiness to seek such evidence in the first place especially after counter-intuitive testimony (see Figure 3).

Third, there was no indication that younger children explored the dolls less than older children because they believed that such exploration was unacceptable. That is, we did not see an increase in younger children’s exploration when they received a prompt implying permission to explore the dolls in the experimenter’s absence. Recall that for half the children in each age group, the experimenter pushed the tray so that the dolls were within easy reach, saying: “I’ll move the dolls a bit closer to you.” This prompt did not influence younger children’s willingness to touch the dolls and it had no impact on younger and older children’s investigation of the dolls. In any case, preschool and elementary school children’s exploration differed primarily in the counter-intuitive

testimony condition. By implication, older children were not generally more bold or exploratory than younger children.

We now consider the implications of the main finding – the age change in children’s exploration when given counter-intuitive testimony. This result adds an important twist to prior research on children’s ability to consider conflicting evidence – specifically, conflicts between what they see and what they are told. As outlined in the introduction, this line of research has shown that when presented with testimonial evidence that conflicts with their perceptions and intuitions, children typically resolve that conflict by deferring to the testimony (e.g., Gelman & Markman, 1986; 1987). Indeed, children will even pass on such counter-perceptual claims to someone else (Jaswal, Lima, & Small, 2009). Our results make the important point that although children are willing to entertain and transmit counter-intuitive claims, their endorsement of such claims need not imply unreflective acceptance of them. Particularly among elementary school children, the counter-intuitive claim prompted empirical exploration.

This result also provides a fresh perspective on the impact of instruction on children’s exploration. Previous research has shown that instruction restricts exploration by reducing the number of hypotheses that children consider (e.g., Bonawitz et al., 2011; Shafto, Goodman, & Frank 2012; Shneidman, Gweon, Schulz, & Woodward, 2016). Our results suggest, however, that for older children whether instruction limits exploration depends on the exact nature of what they are told. When instruction does not conflict with children’s intuitions about what they observe, it may lead them to focus their exploration on a subset of the various possibilities that they would have investigated on their own. This allows children to restrict their exploration in an efficient fashion. By contrast, when

older children are presented with information that conflicts with their intuitions, such information helps them to learn by prompting them to consider possibilities they would not have considered otherwise, thereby increasing their exploration.

A noteworthy implication of the age change in children's exploration following counter-intuitive claims is that children react differently when they observe counter-intuitive phenomena as compared to when they hear counter-intuitive claims. Observable, counter-intuitive phenomena tend to provoke first-hand exploration even in infants (Stahl & Feigenson, 2015) whereas counter-intuitive claims are often accepted without any empirical investigation – at least by preschoolers. By implication, children feel a greater need to investigate discrepancies between what they expect and what they observe than between what they expect and what they are told. This selective pattern of exploration further highlights the privileged and potent role of intentionally communicated information in children's reasoning (Csibra & Gergely, 2009).

Older children's greater exploration following counter-intuitive rather than confirming testimony opens up three questions for future research. First, does this age change apply to children's search for other types of claim? Children are exposed to many counter-intuitive claims. Some of these claims can be checked through observation or experimentation. Other claims cannot. When children are faced with a counter-intuitive claim that they cannot test, are they more likely to query trusted informants and use consensus among these informants to determine whether to accept or reject this claim?

Second, is children's exploration following a counter-intuitive claim influenced by the informant's confidence in her claim? Younger children may be prone to explore more following counter-intuitive claims that are expressed with uncertainty rather than

certainty. In contrast, older children's exploration may be primarily influenced by the nature of the claim – its counter-intuitive nature.

Third, what role does culture play in shaping children's search for information following counter-intuitive claims? We suspect that the age-related differences we observed in our Chinese sample will also be found in other cultures. This is because research confirming an early bias to trust testimony has mostly been conducted in the United States, notwithstanding the socialization of U.S. children towards independence rather deference (Chen & French, 2008; Markus & Kitayana, 1991). Nonetheless, important cultural difference may arise in children's willingness to spontaneously report on their exploration – a tendency that was quite limited in the current sample of Chinese children. This would be an important finding because a greater willingness to report on inconsistent evidence may lead to more child-adult conversation about why such discrepancies exist.

In sum, young children are receptive to claims that defy their perceptions and intuitions. This allows them to quickly acquire many beliefs and practices that they would not be able to learn on their own. However, as this experiment demonstrates, as they get older, children increasingly seize opportunities to evaluate counter-intuitive claims through empirical investigation when such opportunities are available.

Conclusion

Across three studies, this dissertation explored the ability of socially-transmitted, counterintuitive, claims to influence children's beliefs and whether this differed based on children's cultural environment (i.e., China vs. the U.S.). We first discuss similarities between the Chinese and American sample before discussing differences.

The first study asked whether the impact of receiving socially-transmitted, counterintuitive information on children's beliefs was long-lasting rather short-lived. This study revealed that American and Chinese children are willing to consider testimony that conflicts with their intuitions and that this influence of testimony on their beliefs is long-lasting. It continues to influence American and Chinese children's beliefs two weeks after exposure to the testimony. However, the results also suggested that while powerful and long-lasting one exposure to testimony was unlikely to be enough to generate complete belief change. When they were retested 1-to 2-weeks later, American and Chinese children still categorized fewer hybrids based on the perceptual evidence than they had done initially – but more often than they had done immediately following counter-perceptual testimony. By implication, sustained acceptance of counter-perceptual and counter-intuitive claims may require repeated exposure from a single informant or a single exposure from multiple informants to prevent children's beliefs from reverting to their initial beliefs – at least when those beliefs can be reactivated by perceptual evidence.

The second study examined whether American children's exposure to claims that contradict their intuitions triggers their search for evidence. Indeed, some children, particularly older children, sought empirical evidence and questioned their interlocutor

following exposure to a counter-intuitive claim. The third study used a similar task to test whether Chinese children's exposure to claims that contradict their intuitions triggers their search for evidence. Indeed, like American children, older Chinese children more often sought empirical evidence following the experimenter's counter-intuitive claims. In fact, our inclusion of a confirming testimony condition showed that older children's exploration increased when they received counter-intuitive rather than confirming testimony whereas younger children's exploration remained the same. Thus, older children purposefully sought evidence following the counter-intuitive claim. In contrast, younger children "accidentally" encountered this evidence. In both cases, however, if children had explored the dolls they were now more likely to reject the experimenter's counter-intuitive claim than if they had not obtained such evidence.

In sum, converging evidence across these three studies and across two cultures suggests that counter-intuitive testimony may set in motion a process of inquiry that culminates in American and Chinese children's longer-term acceptance or rejection of what they were told. That is, the evidence children uncover whether as a result of their deliberate search for evidence or as a result of random exploration, and their discussion of this evidence with others are all likely to influence what children come to believe and ultimately what they choose to transmit to other people.

As the previous paragraphs make clear, there are important similarities in the manner in which American and Chinese preschoolers responded to the counter-perceptual and counter-intuitive testimony they received. We now turn to differences in American and Chinese preschoolers' memory for, and search for empirical information following, counter-intuitive claims.

We found that the counter-perceptual testimony from the informant identified as a teacher was longer-lasting than the testimony from the informant identified as a mother in the Chinese sample but not in the U.S. sample. When retested, Chinese children were more likely to continue to endorse the counter-perceptual testimony if it had been provided by the informant identified as a teacher rather than by the informant identified as a mother. No such effect was found in the American sample. One interpretation of this finding is that children's socialization influences the amount of epistemic authority they ascribe to a given informant based on that informant's social role. In our case, the high respect placed on teachers in Chinese culture (Li, 2012) as well as parents' desire to see their children succeed in school (Tobin, Hsueh, & Karasawa, 2009) may have led children to be particularly attentive to the testimony provided by the teacher informant. However, an important caveat is that we did not measure children's conception of the knowledge and authority of the informants. Accordingly, more research is needed to secure this interpretation. Nonetheless, it suggests that cultural input about the expertise and authority of particular informants plays a role in children's memory for the counter-intuitive claims they are given.

We found two noteworthy differences between Chinese and American children's endorsement, and selective search for information following, a counter-intuitive claim (i.e., Study 2 vs. Study 3). These difference are somewhat difficult to interpret given differences in the two studies' protocols. However, a few tentative claims can be advanced until a more direct comparison can be made based on identical experimental protocols. First, we found that American children were less likely to endorse the experimenter's counter-intuitive claim than Chinese children (about 65% of the time vs.

about 90% of the time). Secondly, while children in the U.S. and China were both more likely to spontaneously talk about the weight of the dolls upon the experimenter's return if they had gathered empirical evidence contradicting the experimenter's claim, children in the U.S. did so more frequently than children in China (72% vs. 16%). In sum, Chinese children more often deferred to the experimenter than the American children and were less likely to bring up contradictions between what the experimenter had said and what they discovered. These differences are all consistent with the hypothesis outlined in the introduction that American and Chinese children are socialized to advance their own beliefs differently at home and in school when these intuitions contradict the claims of other people. However, a comparison of American and Chinese children's exploration of dolls in the experimenter's absence suggests no difference in Chinese and American children's propensity to seek evidence testing the experimenter's claim when left alone. This is not entirely surprising because children in the U.S. and China may have both felt the need to assuage their uncertainty regarding the experimenter's testimony albeit for different reasons, i.e., "was I right to trust her?" vs. "was I right not to trust her and instead trust myself?"

The difference in Chinese children's behavior relative to American children in a public but not a private setting is consistent with prior research findings. Chinese-American and Chinese children are more likely to endorse counter-perceptual testimony provided by a consensus of multiple informants than European-American children when asked to make these judgments *publicly* (Corriveau et al., 2013; Corriveau & Harris, 2010). However, they do not differ from their Euro-American counterparts when making these judgments *privately* (Corriveau & Harris, 2010) or when asked to transmit such

testimony to someone else (Chan & Tardif, 2013). Moreover, although Chinese-American children are more likely to imitate inefficient actions than Euro-American children when taught by multiple models, they do not differ from Euro-American children when taught by a single model (DiYanni, Corriveau, Kurkul, Nasrini, & Nini, 2015). A plausible explanation for this collection of findings as well as our own is that Chinese and American-Chinese children may be more willing to consider socially provided counter-intuitive and counter-perceptual information in a public setting (perhaps as a means to minimize social conflict) even if they do not believe the testimony they have been given. That is, Chinese children's greater deference likely reflects greater sensitivity to the social context than actual differences in beliefs.

In sum, children are willing to endorse counterintuitive claims and their longer-term acceptance or rejection of such claims is influenced by their subsequent discovery of evidence confirming or disconfirming these claims. This is true of children growing up in two different cultural contexts – the U.S. and China. Despite these similarities, children in these two countries differ in the extent to which they are willing to publicly voice agreement with counter-intuitive claims. In the U.S. where children are encouraged to voice independent opinions they are more likely to voice disagreement with such claims than in China where such behavior is less encouraged. Thus, the experiments reveal important but rather subtle differences in how children approach claims that contradict their perceptions and intuitions: children vary in their initial consideration of these claims but not necessarily in their long-term acceptance of them – acceptance being determined by children's search for additional information regarding the counter-intuitive claim and how long they have been exposed to that claim.

In the sections that follow I briefly discuss the implication of these three studies for our understanding of children's cognitive development. I then turn to implications for future research. Specifically, I discuss possible relations between children's socio-communicative environments, children's propensity to seek information through questions, observation, and experimentation. Then, I discuss the possible role of executive function skills for children's ability to learn from socially transmitted and empirical information. I conclude by discussing the implications of these studies for early childhood education.

Implications for children's cognitive development

These three studies further our understanding of children's cognitive development in two ways. First, Studies 1A and 1B demonstrate that the acquisition of counterintuitive claims is a protracted process that is best understood using methods that measure children's acceptance of such claims over time. Second, Studies 2 and 3 point to connections between children's search for information through exploration and question-asking –two means of acquiring information that have thus far been studied independently of one another.

As children age they increasingly encounter and come to believe in claims that defy their first-hand perceptions and their intuitions about the world (Lane & Harris, 2014). This dissertation advances our understanding of children's acquisition of such concepts by demonstrating that coming to believe counterintuitive claims is a protracted process that extends beyond exposure to a single claim. This conclusion should spur future work to move beyond measures of children's endorsement of counter-intuitive claims to measures of children's engagement with such claims over time. For example, in their

review of the literature on the development of children's counterintuitive concepts, Lane and Harris (2015) identify a number of factors that may influence children acceptance of these concepts (e.g., qualities of informants, qualities of the information, qualities of the context). The methods used in Studies 1A and 1B, Study 2, and Study 3 provide means to extend our understanding of these factors. For example, Lane and Harris (2015) found that when presented with counterintuitive claims, children between 3- and 8-years-old were more trusting of claims provided by experts with relevant (as opposed to irrelevant) expertise. However, how strong is this influence of expertise on children's endorsement of these claims? Are counterintuitive claims provided by an expert better remembered over time than those provided by a non-expert? Are such claims less likely to be tested by children if they are given the chance to do so? In other words, what is the depth of children's acceptance of these claims?

Studies 2 and 3 suggest interesting relations between children's questioning and exploratory play. Developmental research has shown that infants and young children are motivated to understand causal relationships and will investigate following their *observation* of novel and unexpected phenomena (Schulz, 2009; Stahl & Feigenson, 2015). Moreover, as Study 2 and 3 demonstrated, young children will also seek empirical evidence and discuss this evidence with an informant following the receipt of unexpected *testimony*. However, young children's exploration of the world is not restricted to empirical inquiry. Children ask many questions often with the purpose of obtaining information. Indeed, children will generate their own explanation or request new explanations following the non-provision of explanations or the provision of non-explanations (Frazier, Wellman, & Gelman, 2009). Thus, children use their observations

and the explanations they receive as fodder for inquiry – whether that inquiry involves querying the world or other people. The fact that these two modes of inquiry are triggered by similar antecedents (i.e., unexpected observations or unexpected claims) suggests a common psychological mechanism and suggests that children’s exploration and questioning mirror and influence each other during early childhood. However, because children’s physical exploration of the world and their exploration through questions have so far been investigated separately we do not know how these two forms of exploration are related and how they influence one another. For example, how do children’s early physical exploration and their caregiver’s response shape their propensity to seek additional information? This is an important goal for future research.

Relations between children’s socio-communicative environments and children’s propensity to seek information

Very little research has examined individual differences in children’s learning from exploration and questions. For example, Legare (2012) finds that children’s provision of a mechanistic (i.e., it’s broken) rather than a category-based explanation (i.e., it must not be a blicket) predicts children’s search for evidence following their observation of a broken blicket, e.g., their engagement in hypothesis testing behaviors such as repeatedly placing the blicket on the machine. However, Legare (2012) did not explore what led children to provide these different types of explanations. Similarly, other studies of children’s exploratory and explanatory abilities do not explore individual differences in children’s engagement in exploratory play and in their provision of explanations.

Differences in children's socio-communicative environments may partially explain individual variation in children's propensity to engage in empirical exploration and verbal questioning as well as their propensity to generate and evaluate the explanations they are given. That is, one way in which children's exploration and questioning might be similarly influenced and might influence one another is through the responses that these acts of inquiry generate from caregivers. Indeed, the amount and kind of talk children receive varies markedly within and across social-class (Hart & Risley, 1992) with important implications for children's language and cognitive development (Hoff & Naigles, 2002; Huttenlocher et al., 2010; Rowe, Leech, Cabrera, 2016).

Caregivers differ in the amount of wh-questions they produce (who, what, where, when, why, how) (Rowe, Coker, Pan, 2004). Parents from higher social classes ask more questions to their children (e.g., Farran & Haskins, 1980; Hart & Risley, 1995; Heath, 1983; Rowe, 2008) and tend to provide more informative replies to their children's questions (Corriveau, 2016). These differences in input are likely to have important implications for the development of children's information seeking. For example, Tizard and Hughes (1984) reported that 4-year-old UK children from higher social classes asked more questions than their lower class peers (see also, McCarthy, 1930). However, it is currently unclear how children's socio-communicative environments influence children's information seeking behaviors. Seeking additional information requires children to identify that they need information, to formulate a question or hypothesis, and then to ask that question or to test that hypothesis. Each of these steps may be influenced by children's socio-communicative environments.

First, children's socio-communicative environments might influence children's search for evidence by influencing children's propensity to monitor their own knowledge. Growing up in a family that often asks for justifications may lead children to internalize this pattern of thought. Internalizing the need to have justifications for their beliefs may lead children to more often reflect on what they know and are told and this may lead them to more frequently identify gaps in their own knowledge and therefore to seek information to address these gaps. Indeed, there is great variability in children and adults' propensity to engage in self-explanation (Rittle-Johnson, 2006) and engaging in self-explanation facilitates learning (Aleven & Koedinger, 2002; Pine & Siegler, 2003; Wong, Lawson, & Keeves, 2002). Thus, one possible pathway to variability in children's information seeking is that differences in the frequency with which children are asked to justify their beliefs (i.e. are asked justification questions or have their explanations corrected) lead children to monitor their understanding more closely which in turn leads them to seek information more often. This greater sensitivity to discrepancies in their own knowledge should influence both children's questioning of people and their empirical investigations.

Second, children's socio-communicative environments may influence their children's ability to ask appropriate questions by shaping their understanding of the evidentiary constraints in a particular domain of knowledge. Understanding the boundaries of a particular domain of knowledge may help children reduce the number of explanations they consider and thus help them ask more pointed questions. Children's explanations suggest that they understand that phenomena in different domains require different kinds of causal explanations. By 2.5-years-old, children spontaneously provide

explanations about various entities (e.g., persons, animals, objects) using different explanatory modes (i.e., modes that use different causal mechanisms: physical, psychological, social-conventional, biological). Importantly, children's explanations pair entities with explanatory modes in a manner that suggests constrained yet flexible causal reasoning (Callanan & Oakes, 1992; Hickling & Wellman, 2001; Wellman, Hickling, & Schult, 1997). For example, children explain the actions of individuals using psychological explanations but use physical explanations to explain an object's movement. Similarly, between 5- and 9-years-old children increasingly understand circumstances where evidence is helpful in resolving conflicts and where it is not. Wainyrb and colleagues (2004) asked 5-, 7-, and 9-year-olds to consider disagreements between two people over a matter of fact, a moral issue, an issue of taste, and an ambiguous causal issue. They asked children if only one or both of the individuals could be right in their conflicting claims. At all three age groups, children almost always said that only one person could be right when debating a matter of fact or a moral issue. However, with respect to an issue of taste or an ambiguous causal issue, children more often stated that both people could be right. This pattern of reasoning which was already presents among 5-year-olds increased in frequency among 7- and 9-years old. An important implication of these findings is that children understand how standards of evidence (i.e., the kinds of explanations and evidence that are required to adjudicate a claims) vary across domains – an important skill for asking appropriate questions and for generating and evaluating explanations.

Children's developing understanding of the appropriateness of causal mechanisms for a given domains and their understanding of the role of evidence in resolving conflicts

is presumably acquired through children's exposure to, and engagement in, conversations and may play an important role in children's information seeking by allowing children to identify the *kinds* of evidence that are required in different domains. That is, another possible pathway to variability in children's information seeking is that children's socio-communicative environments shape children's understanding of the evidentiary constraints of different domains of knowledge and this may help children generate more specific questions whose answer are more likely to help them learn. Understanding the evidentiary constraints of various domains is likely to be easier to observe in children's questions than in their empirical investigations. Nonetheless, just as children's questions may be more precise, children's empirical investigations may also be more precise. For example, children with a better understanding of the kind of evidence required by a particular problem may be quicker to obtain that information (i.e., may gather it in fewer steps).

Finally, children's socio-communicative environments, particularly their caregivers' positive response to their exploration and queries might nurture in children an expectation that seeking information from others is a good strategy for learning. This may in turn influence children's decision to ask rather than keep to themselves the questions and explanations they have formulated. Consider a 12-month-old infant who is exploring puzzle blocks on her own. She turns to her mother and expresses puzzlement at where to place one of the puzzle pieces. Or consider the case of a 3-year-old who asks her mother questions about the cause of a particular event. Does the tendency of a caregiver to respond or ignore these expressions of uncertainty and these requests for information influence a child's information seeking in later years? Moreover, does the quality of the

caregiver's response matter? Thus, another possible pathway to variability in children's information seeking is through children's expectation about the usefulness of seeking information – an expectation based on their prior interactions with familiar caregivers. Children who frequently receive informative replies may be more likely to seek information than children who do not receive informative replies (or any replies). We might expect that the expectation that others will provide informative replies might influence the propensity of children's verbal questions but not their independent empirical investigations.

So far I have discussed three non-exclusive mechanisms through which differences in children's socio-communicative environments might influence children's search for information: influences on children's self-monitoring, influences on children's ability to ask targeted question, and influences on children's expectation that they will receive an informative reply. These three influences are likely to exist across various cultures and communities. However, it is also the case that communities differ widely in their beliefs about how children should interact with adults (Gauvain, Munroe, & Beebe, 2013; Heath, 1983). Specifically, many cultures expect children to be deferential and this may lead children to fewer questions even if children across these different cultures do not differ in their ability and propensity to formulate questions and explanations. Rowe (2008) showed that parents' beliefs about child development influences the frequency of parents' child-directed speech and in turn children's language development. Similarly, parents' beliefs about the importance of allowing children to express their own opinions and to ask questions rather than to be deferential may influence how parents respond to their children's questions and consequently influence the development of children's

information seeking through the pathways identified above. Preliminary evidence for this claim comes from a comparison of Study 2 and Study 3. American children very frequently discussed the results of their exploration with the experimenter (72%) while Chinese children rarely did so (16%). Given the documented cultural differences in deference to authority between the US and China (Chen & French, 2008; Markus & Kitayama, 1991), this suggests that children's propensity to share what they discover when it conflicts with what they are taught may be shaped by their home environment. Additional suggestive evidence of this relationship can be inferred from Gauvain, Munroe and Beebe (2013)'s analysis of children's conversations in Belize, Kenya, Nepal and Samoa –countries where adults expect children to be deferential. The authors found that children's information-seeking questions made up about one-tenth of the remarks that children made, a proportion similar to that observed by Chouinard (2007) in the U.S. However, in these communities, unlike in the US, children very rarely asked how or why questions – questions one might expect following an adult's provision of information and which might be interpreted as doubting or probing that explanation.

In sum, it will be important for future work to examine how different manifestations of children's search for information (i.e., their physical exploration of the world and their questioning) are related to and influence one another. One likely mechanism influencing both of these forms of information seeking is children's socio-communicative environment – caregivers' propensity to provide, request, and evaluate explanations in their conversations with their children, a propensity that is likely to be rooted in parents' tacit theories of child development and in their child development goals.

The role of executive function skills and children's ability to learn from inquiry

The previous section discussed three pathways through which children's socio-communicative environments might influence their search for information. In this section, I turn to children's ability to learn from their inquiries and specifically to the role that children's executive function skills (EF) might play in this process. Executive functions skills refer to three sets of cognitive processes: inhibition (self-control – behavioral inhibition, and interference control – selective attention and cognitive inhibition), working memory, and cognitive flexibility (the ability to switch back and forth between different ways of thinking about phenomena) (Diamond, 2013). These three components of EF are all likely to be involved in children's ability to learn from the information they gather through questions or through empirical investigations. This is because to learn from an explanation about phenomena or from their exploration of a particularly phenomenon, children must inhibit their current beliefs (cognitive inhibition), evaluate the extent to which a new explanation accounts for the data at hand relative to their prior beliefs (working memory and cognitive flexibility), and update their mental model of the phenomena under consideration – all skills that develop rapidly during the preschool years. Indeed, recent experimental work suggests that working memory plays an important role in children's ability to learn from an explanation (Bascandziev, Powell, Harris, & Carey, 2016). Bascandziev et al. (2016) asked children to find a ball that was dropped down an opaque curved tube. When asked to perform this task, two- and three-year-old children exhibit a gravity bias. They tend to search for the ball directly below the place where the ball was dropped rather than at the bottom of the tube into which the ball was dropped. The authors found that children could overcome this bias if provided with

an explanation about the causal role of the tube in constraining the downward path of the ball. Importantly, children's ability to use this explanation to make correct predictions about the location of the ball on future trials was predicted by higher scores on a working memory task, controlling for age, gender, and receptive vocabulary. This result suggests that while it is important to understand how children develop the ability to question the world and the people around them and how such abilities can be nurtured, it is equally important to study the factors that influence children's ability to learn from the information they gather. It is possible that EF skills play an important role in children's ability to generate questions and hypotheses as well as in their ability to learn from the information they have gathered. However, it may also be the case that EF skills plays a greater role in one of these two processes or that different facets of EF are important for generating questions, evaluating the data that are obtained as a result of these questions, and updating existing beliefs to incorporate these new conclusions in mental representations of the world. Moreover, EF skills may be particularly important when children are learning about concepts that contradict their everyday experiences and intuitions. For example, could differences in EF explain variability in children's memory for the counterintuitive testimony they received in Study 1? Or for their propensity to seek evidence and learn from that evidence in Study 2?

Implications for early childhood education

The line of research pursued in this dissertation and its results have important implications for educational practice. They demonstrate that children are prepared to learn from teachers in domains where what they are taught conflicts with what they know or see and that such instruction is memorable – children encode this information into long

term memory. In addition, it appears that some children's willingness to learn from their teachers is bounded—particularly in older children. In Studies 2 and 3, we found that some children will question what they are told by seeking evidence and discussing that evidence with their informant. This spontaneous exploration of instruction may be harnessed in schools to improve learning outcomes. Thus, an important implication of that finding is that instruction does not always come at the expense of children's exploration and curiosity. Debates in education often center on the value of direct-instruction versus more student centered method of instruction. However, the current results suggest that instruction and exploration are not enemies and may even work together to support students' learning. More research needs to be done to better understand the conditions under which instruction and exploration interact to support learning. However, the current study demonstrates that direct instruction and exploration can be used together to promote children's critical thinking about what they are told and what they discover.

In the previous section, we discussed the possible role of children's socio-communicative environment in the home on the development of their information seeking skills. If children's socio-communicative environment at home influences their development so must the socio-communicative environment of their schools. Children spend a lot of time in schools and, as a result, schools play an important role in children's development. Unfortunately, not much discussion occurs in classroom. In a study of the effects of instructional strategies on achievement in eighth grade Social Studies and English classes, Gamoran and Nystrand (1991) found that authentic discussion that allowed students to contribute their own thoughts and ideas occurred on average less than

one minute per day. Similarly, in a follow-up study of eighth- and ninth-grade English and Social Studies classrooms, they found that; “discussion in English took 50 seconds per class in eighth grade and less than 15 seconds in ninth grade. Average time for discussion in Social Studies was 42 seconds in eighth grade and 31.2 seconds in ninth grade” (Nystrand, Wu, Gamoran, Zeiser, & Long, 2003, p. 178). Moreover, in these discussions, teachers most often requested memorized facts from students rather than press or pull for elaborated conversation. Thus, increasing discussion and debate, particularly in earlier grades and preschool through activities that combine exploration and instruction may be an important lever for developing children’s inquiry and explanatory skills. Indeed, recent work demonstrates that increasing discussion and debate in middle school classrooms influences children’s language development (Lawrence, Crosson, Paré-Blagoev, & Snow, 2015).

Final thoughts

How do children make sense of the world? One influential view is that young children are “little scientists” who independently make inferences and test hypotheses as they interact with the world around them. Another influential view is that children are profoundly shaped by the people around them. These two perspectives complement each other and suggest that children may indeed be more like little scientist than has been originally argued. That is, young children are not independently making sense of the world through exploration. Instead, just like “real” scientists, their exploration of the world is shaped by their community and what they learn, in turn, shapes their interaction with their community and changes the community itself.

References

- Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, 26, 147–179.
- Bascandziev, I., Powell, L. J., Harris, P. L., & Carey, S. (2016). A role for executive functions in explanatory understanding of the physical world. *Cognitive Development*, 39, 71-85.
- Bernard, S., Harris, P., Terrier, N., & Clément, F. (2015). Children weigh the number of informants and perceptual uncertainty when identifying objects. *Journal of Experimental Child Psychology*, 136, 70–81. doi: 10.1016/j.jecp.2015.03.009
- Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, 120, 322-330. doi:10.1016/j.cognition.2010.10.001
- Callanan, Maureen A., and Lisa M. Oakes. (1992). Preschoolers' questions and parents' explanations: Causal thinking in everyday activity. *Cognitive Development*, 7, 213-233.
- Chan, C. C. Y., & Tardif, T. (2013). Knowing better: The role of prior knowledge and culture in trust in testimony. *Developmental Psychology*, 49, 591–601. doi: 10.1037/a0031336
- Chen, X., & French, D. C. (2008). Children's social competence in cultural context. *Annual Review of Psychology*, 59, 591–616. doi: 10.1146/annurev.psych.59.103006.093606

- Chouinard, M. (2007). Children's questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development*. Serial No. 286, 72, No. 1
- Clément, F., Koenig, M. A., & Harris, P. L. (2004). The ontogenesis of trust. *Mind & Language*, 19, 360–379. doi: 10.1111/j.0268-1064.2004.00263.x
- Corriveau, K.H. (2016). *Individual differences in children's use of questions to acquire knowledge*. Radcliffe Institute Workshop on Questions, Cambridge, MA
- Corriveau, K. H., & Harris, P. L. (2009a). Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information. *Developmental Science*, 12, 188–193. doi: 10.1111/j.1467-7687.2008.00763.x
- Corriveau, K. H., & Harris, P. L. (2009b). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science*, 12, 426–437. doi: 10.1111/j.1467-7687.2008.00792.x
- Corriveau, K. H., & Harris, P. L. (2010). Preschoolers (sometimes) defer to the majority when making simple perceptual judgments. *Developmental Psychology*, 26, 437–445. doi: 0.1037/a0017553
- Corriveau, K. H., Harris, P. L., Meins, E., Fernyhough, C., Arnott, B., Elliott, L., & De Rosnay, M. (2009c). Young children's trust in their mother's claims: Longitudinal links with attachment security in infancy. *Child Development*, 80, 750–761. doi: 10.1111/j.1467-8624.2009.01295.x
- Corriveau, K. H., Kim, E., Song, G. & Harris, P. L. (2013). Young children's deference to a majority varies by culture. *Journal of Cognition and Culture*, 13, 367–381. doi: 10.1163/15685373-12342099

- Diamond, A. (2013). Executive functions. *Annual review of psychology*, 64, 135-168.
- DiYanni, C., Corriveau, K. H., Kurkul, K., Nasrini, J., & Nini, N. (2015). The role of consensus and culture in the imitation of questionable actions. *Journal of Experimental Child Psychology*, 137, 99-110. doi: 10.1016/j.jecp.2015.04.004
- Farran, D. C., & Haskins, R. (1980). Reciprocal influence in the social interactions of mothers and three-year-old children from different socioeconomic backgrounds. *Child Development*, 51, 780-791.
- Flavell, J. H., Green, F. L., & Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. *Monographs of the Society for Research in Child Development*, 51 (1, Serial No. 212).
- Frazier, B. N., Gelman, S. A., & Wellman, H. M. (2009). Preschoolers' search for explanatory information within adult-child conversation. *Child development*, 80, 1592-1611. doi: 10.1111/j.1467-8624.2009.01356.x
- Gamoran A., & Nystrand M. (1991). Background on instructional effects on achievement in eighth-grade English and social studies. *Journal of Research on Adolescence*, 1, 277-300.
- Gauvain, M., Munroe R. L. & Beebe (2013). Children's questions in cross-cultural perspective: A four-culture study. *Journal of Cross-Cultural Psychology*, 44, 1148-1165.
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, 23, 183-209. doi:10.1016/0010-0277(86)90034-X

- Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development*, 58, 1532-1541. doi: 10.2307/1130693
- Harkness, S., Mavridis, C.J., Liu, J.L., Super, C.M. (2015). Parental Ethnotheories and the Development of Family Relationships in Early and Middle Childhood. In Jensen, L.A. (Ed.). *The Oxford Handbook of Human Development and Culture*. Oxford, UK: Oxford University Press.
- Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Cambridge, MA: Belknap Press/Harvard University Press.
- Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development*, 77, 505–524. doi: 10.1111/j.1467-8624.2006.00886.x
- Hart, B. & Risley, T. (1992). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology*, 28, 1096-1105.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD, US: Paul H Brookes Publishing.
- Heath, S. B. (1983). *Ways with words: Language, life and work in communities and classrooms*. Cambridge, UK: Cambridge University Press.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and brain sciences*, 33, 61-83.

- Hickling, A. K., & Wellman, H. M. (2001). The emergence of children's causal explanations and theories: Evidence from everyday conversation. *Developmental Psychology*, 37, 668.
- Hoff, E. & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, 73, 418-433.
- Huttenlocher, J., Waterfall, H.R., Vasilyeva, M., Vevea, J.L. & Hedges, L.V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, 61, 343-365.
- Jaswal, V. K. (2004). Don't believe everything you hear: Preschoolers' sensitivity to speaker intent in category induction. *Child Development*, 75, 1871–1885. doi: 10.1111/j.1467-8624.2004.00822.x
- Jaswal, V. K., Lima, O. K., & Small, J. E. (2009). Compliance, conversion, and category induction. *Journal of Experimental Child Psychology*, 102, 182–195. doi: 10.1016/j.jecp.2008.04.006
- Jaswal, V. K., & Markman, E. M. (2007). Looks aren't everything: 24-month-olds' willingness to accept unexpected labels. *Journal of Cognition and Development*, 8, 93–111. doi: 10.1080/15248370709336995
- Koenig, M. A., & Echols, C. H. (2003). Infants' understanding of false labeling events: The referential roles of words and the speakers who use them. *Cognition*, 87, 179–208. doi: 10.1016/S0010-0277(03)00002-7
- Lancy, D. R, Gaskins, S. & Bock, J. (eds) (2009). *The anthropology of learning in childhood*. Lanham, MD: Altamira Press.

- Lane, J. D., & Harris, P. L. (2014). Confronting, representing, and believing counterintuitive concepts navigating the natural and the supernatural. *Perspectives on Psychological Science*, 9, 144-160. doi: 10.1177/1745691613518078
- Lane, J. D., & Harris, P. L. (2015). The roles of intuition and informants' expertise in children's epistemic trust. *Child Development*, 86, 919–926. doi: 10.1111/cdev.12324
- Lane, J. D., Harris, P. L., Gelman, S. A., & Wellman, H. M. (2014). More than meets the eye: Young children's trust in claims that defy their perceptions. *Developmental Psychology*, 50, 865-871. doi: 10.1037/a0034291
- Lane, J. D., Ronfard, S., Francioli, S. P., & Harris, P. L. (2016). Children's imagination and belief: Prone to flights of fancy or grounded in reality?. *Cognition*, 152, 127-140. doi:10.1016/j.cognition.2016.03.022
- Lawrence, J. F., Crosson, A. C., Paré-Blagoev, E. J., & Snow, C. E. (2015). Word generation randomized trial discussion mediates the impact of program treatment on academic word learning. *American Educational Research Journal*, 52, 750-786
- Legare, C. H. (2012). Exploring explanation: Explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child development*, 83, 173-185. doi: 10.1111/j.1467-8624.2011.01691.x
- Legare, C. H. (2014). The contributions of explanation and exploration to children's scientific reasoning. *Child Development Perspectives*, 8, 101-106. doi: 10.1111/cdep.12070

- Legare, C.H., Gelman, S.A., & Wellman, H.M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development, 81*, 929-944. doi: 10.1111/j.1467-8624.2010.01443.x
- Li, J. (2012). *Cultural foundations of learning: East and West*. Cambridge University Press.
- Lopez-Mobilia, G., & Woolley, J. D. (2016). Interactions between knowledge and testimony in children's reality-status judgments. *Journal of Cognition and Development*. Advanced online publication. doi: 10.1080/15248372.2015.1061529
- Markus, H., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224-253.
- McCarthy, D. A. (1930). *The language development of the preschool child*. Minneapolis, MI: University of Minnesota Press.
- Mounoud, P., & Bower, T.G.R. (1974). Conservation of weight in infants. *Cognition, 3*, 29-40.
- Nielsen, M., & Haun, D. (2016). Why developmental psychology is incomplete without comparative and cross-cultural perspectives. *Philosophical Transactions of the Royal Society B, 371*, 20150071
- Pea, R. D. (1982). Origins of verbal logic: Spontaneous denials by two- and three-year olds. *Journal of Child Language, 9*, 597-626. doi: 10.1017/S0305000900004931
- Pine, K., J., & Siegler, R., S. (2003). *The role of explanatory activity in increasing the generality of thinking*. Paper presented at the biennial meeting of the Society for Research in Child Development, Tampa, FL.

- Rittle-Johnson, B. (2006). Promoting transfer: The effects of direct instruction and self-explanation. *Child Development*, 77, 1–15.
- Robinson, E. J., Champion, H., & Mitchell, P. (1999). Children's ability to infer utterance veracity from speaker informedness. *Developmental Psychology*, 35, 535–546.
- Robinson, E. J., Mitchell, P., & Nye, R. M. (1995). Young children's treatment of utterances as unreliable sources of knowledge. *Journal of Child Language*, 22, 663–685. doi: 10.1017/S0305000900009995
- Rogoff, B. (2003). *The cultural nature of human development*. New York, NY: Oxford University Press.
- Rowe, M.L. Coker, D. & Pan, B.A. (2004). A comparison of fathers' and mothers' talk to toddlers in low-income families. *Social Development*, 13, 278-291.
- Rowe, M. L. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, 35, 185-205
- Rowe, M. L., Leech, K. A., & Cabrera, N. (2016). Going beyond input quantity: wh-questions matter for toddlers' language and cognitive development. *Cognitive science*. Online first publication. doi: 10.1111/cogs.12349
- Sabbagh, M. A., & Shafman, D. (2009). How children block learning from ignorant speakers. *Cognition*, 112, 415-422. doi: 10.1016/j.cognition.2009.06.005
- Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, 16, 382-389.
doi:10.1016/j.tics.2012.06.004

- Shafto, P., Goodman, N.D. & Frank, M.C. (2012). Learning from others: The consequences of psychological reasoning for human learning. *Perspectives on Psychological Science*, 7, 341-351. doi: 10.1177/1745691612448481
- Shneidman, L., Gweon, H., Schulz, L. E., & Woodward, A.L. (2016). Learning from others and spontaneous exploration: A cross-cultural investigation. *Child Development*, 78, 723-735.
- Shtulman, A. (2009). The development of possibility judgment within and across domains. *Cognitive Development*, 24, 293-309. doi: 10.1016/j.cogdev.2008.12.006
- Shtulman, A., & Carey, S. (2007). Improbable or impossible? How children reason about the possibility of extraordinary events. *Child Development*, 78, 1015-1032.
- Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, 348(6230), 91-94.
- Tobin, J., Hsueh, Y., & Karasawa, M. (2009). *Preschool in three cultures revisited: China, Japan, and the United States*. Chicago: University of Chicago Press.
- Van Schijndel, T. J. P., Visser, I., Van Bers, B. M. C. W., & Raijmakers, M. E. J. (2015). Preschoolers perform more informative experiments after observing theory-violating evidence. *Journal of Experimental Child Psychology*, 131, 104 – 119.
- Wainyrb, C., Shaw, L.A., Langley, M., Cottam, K. & Lewis, R. (2004). Children's thinking about diversity of belief in the early school years: Judgments of relativism, tolerance, and disagreeing persons. *Child Development*, 65, 687-703.
- Wellman, H. M., Hickling, A. K., & Schult, C. A. (1997). Young children's psychological, physical, and biological explanations. *New Directions for Child*

and Adolescent Development, 75, 7-26.

Wang, J., Elicker, J., McMullen, M., & Mao, S. (2008). Chinese and American preschool teachers' beliefs about early childhood curriculum. *Early Child Development and Care*, 178, 227-249.

Wong, R., M., F., Lawson, M., J., & Keeves, J. (2002). The effects of self-explanation training on students' problem solving in high-school mathematics. *Learning and Instruction*, 12, 233–262.

Ziv, M., & Frye, D. (2004). Children's understanding of teaching: the role of knowledge and belief. *Cognitive Development*, 19, 457–477. doi:
10.1016/j.cogdev.2004.09.002









Appendix A

A pre-test was conducted with a sample of 16 younger ($M = 3.97$, $SD = .53$, $R = 2.94 - 4.67$) and 16 older children ($M = 5.70$, $SD = .66$, $R = 4.97 - 7.06$) to assess children's perception of each hybrid. For each child we presented 10 hybrids (75%-25% from Jaswal & Markman, 2007): 5 animals (cat-dog, horse-cow, squirrel-rabbit, bear-pig, bird-fish) and 5 objects (spoon-key, car-shoe, button-ball, hat-cup, pen-toothbrush). The experimenter presented the ten hybrids one at a time using laminated pictures and told children: "I am going to show you a picture, do you think this is a [75% label] or a [25% label]?" The order of the labels was counterbalanced across children. The experimenter randomized the order of presentation of the hybrids by shuffling the pictures before testing each child.

Potential age differences were assessed using a Kruskal-Wallis one-way analysis of variance. Two items had marginally significant main effects of age group and were therefore deemed inappropriate for the experiment (button-ball: $\chi^2(1, N=32) = 3.36$, $p = .067$; horse-cow: $\chi^2(1, N=32) = 3.74$, $p = .053$). We combined the remaining eight items into two sets. Each set contained two animals and two objects: set 1 (*cat-dog* = 84%, *squirrel-rabbit* = 75%, *hat-cup* = 94%, *pen-toothbrush* = 41%); set 2 (*bird-fish* = 88%, *bear-pig* = 66%, *car-shoe* = 84%, *spoon-key* = 63%). Kruskal-Wallis one-way analysis of variance revealed no age-group differences for set 1 or set 2. To confirm that these sets did not differ from each other, we added the four items from each set and compared them using a paired-samples t-test. This confirmed that set 1 ($M = 2.94$, $SD = .80$) and set 2 ($M = 3.00$, $SD = .88$) did not differ from each other, $t(31) = -.39$, *ns.* We also tested for differences between animals and objects. We combined all four animal items together to

create an animal composite and combined all four object items together to create an object composite. Kruskal-Wallis one-way analysis of variance revealed no age-group differences for the object composite or the animal composite. We compared the animal composite ($M = 3.13$, $SD = .87$) and object composite ($M = 2.81$, $SD = 1.00$) using a paired-samples t-test. These two composites did not differ from each other, $t(31) = 1.43$, *ns*.

Below, we display the final stimuli and accompanying reasoning questions associated with each hybrid for set 1 and set 2. The hybrids combined 75% of one animal or object and 25% from another. The first mentioned name (*italicized*) contributed 75% of the perceptual evidence for the hybrid

Set	Hybrid	Picture	Reasoning Question
1	<i>Cat-dog</i>		Do you think this animal barks or meows?
1	<i>Squirrell-rabitt</i>		Do you think this animal eats nuts or carrots?
1	<i>Hat-cup</i>		Do you think people use it to drink or wear it on their head?
1	<i>Pen-toothbrush</i>		Do you think people use it to brush their teeth or to write?
2	<i>Bird-fish</i>		Do you think this animal flies or swims?
2	<i>Bear-pig</i>		Do you think this animal growls or oinks?
2	<i>Car-shoe</i>		Do you think people drive it or wear it on their feet?
2	<i>Spoon-key</i>		Do you think people use it to open doors or to eat?

Appendix B

A pre-test was conducted with a sample of 16 younger (11 females, $M = 4.48$, $SD = .39$, $R = 3.78 - 4.99$) and 16 older children (9 females, $M = 6.27$, $SD = .39$, $R = 5.63 - 6.81$) to assess children's perception of each hybrid. For each child we presented 10 hybrids (75%-25% from Jaswal & Markman, 2007): 5 animals (cat-dog, horse-cow, squirrel-rabbit, bear-pig, bird-fish) and 5 objects (spoon-key, car-shoe, button-ball, hat-cup, pen-toothbrush). The experimenter presented the ten hybrids one at a time using laminated pictures and told children: "I am going to show you a picture, do you think this is a [75% label] or a [25% label]?" The order of the labels was counterbalanced across children. The experimenter randomized the order of presentation of the hybrids by shuffling the pictures before testing each child.

Potential age differences were assessed using a Kruskal-Wallis one-way analysis of variance. One item had a marginally significant main effect of age group and was therefore deemed inappropriate for the experiment (pen-toothbrush: $\chi^2(1, N=32) = 3.35$, $p = .067$). We selected eight of the remaining nine items to create two sets of four items each. Each set contained two animals and two objects: set 1 (*cat-dog* = 81%, *bear-pig* = 72%, *hat-cup* = 100%, *button-ball* = 31%); set 2 (*bird-fish* = 91%, *squirrel-rabbit* = 47%, *car-shoe* = 91%, *spoon-key* = 69%). Kruskal-Wallis one-way analysis of variance revealed no age-group differences for set 1 or set 2. To confirm that these sets did not differ from each other, we added the four items from each set and compared them using a paired-samples t-test. This confirmed that set 1 ($M = 2.97$, $SD = .69$) and set 2 ($M = 2.75$, $SD = .84$) did not differ from each other, $t(31) = -.39$, *ns*. We also tested for differences between animals and objects. We combined all four animal items together to create an

animal composite and combined all four object items together to create an object composite. Kruskal-Wallis one-way analysis of variance revealed no age-group differences for the object composite or the animal composite. We compared the animal composite ($M = 2.91$, $SD = .89$) and object composite ($M = 2.81$, $SD = .82$) using a paired-samples t-test. These two composites did not differ from each other, $t(31) = 1.56$, *ns*.

Below, we display the final stimuli and accompanying reasoning questions associated with each hybrid for set 1 and set 2. The hybrids combined 75% of one animal or object and 25% from another. The first mentioned name (*italicized*) contributed 75% of the perceptual evidence for the hybrid.

Set	Hybrid	Picture	Reasoning Question
1	<i>Cat-dog</i>		Do you think this animal barks or meows?
1	<i>Squirrell-rabitt</i>		Do you think this animal eats nuts or carrots?
1	<i>Hat-cup</i>		Do you think people use it to drink or wear it on their head?
1	<i>Button-Ball</i>		Do you think people use it to play or to put on their shirt?
2	<i>Bird-fish</i>		Do you think this animal flies or swims?
2	<i>Bear-pig</i>		Do you think this animal growls or oinks?
2	<i>Car-shoe</i>		Do you think people drive it or wear it on their feet?
2	<i>Spoon-key</i>		Do you think people use it to open doors or to eat?

Appendix C

Analyses of whether children picked up the smallest and the biggest doll *concurrently*.

A significant interaction between Age Group and Testimony Type emerged (Table C1). Preschoolers' exploration did not differ by type of testimony, GLH Test: $\chi^2(1) = .72, p > .25$. However, significantly more elementary school children explored following counter-intuitive than confirming testimony, GLH Test: $\chi^2(1) = 6.11, p = .013$. Thus, more elementary school children than preschool children explored following counter-intuitive testimony, GLH Test: $\chi^2(1) = 7.49, p = .006$, whereas there was no age difference in exploration following confirming testimony, GLH Test: $\chi^2(1) = .99, p > .25$. There were no further interactions.

Table C1. Logistic regression model comparing whether children picked up the biggest and smallest dolls as a function of the type of testimony children received, whether they received a prompt, and their age.

	Model 2		
	Odds-Ratios	z scores	95% CI
Counter-Intuitive Testimony	.37	.85	.04, 3.70
Elementary	2.03	.99	.50, 8.19
Prompt	1.93	1.54	.84, 4.47
Counter-Intuitive X Elementary	8.81	1.71~	.71, 106.44
Constant	.05*	4.49***	.01, .18
X²		23.19***	
Model df		4	
-2 Log Likelihood		149.32	

~ $p = .087$; * $p < .05$; ** $p = .011$; *** $p < .001$. *Note.* $n = 200$.

Appendix D

Children's decision to explore was unrelated to whether they endorsed or rejected the experimenter's testimony.

Whether children endorsed the experimenter's testimony that smallest = heaviest or stuck to their initial intuition that biggest = heaviest was unrelated to their decision to pick up the biggest and smallest doll during the experimenter's absence, 65.48% vs. 58.33%, $\chi^2(1, n = 96) = 0.32, p = .63$. This was true for preschool and elementary school children, $\chi^2(1, n = 38) = 0.97, p = .32$, $\chi^2(1, n = 58) = 0.23, p = .63$, respectively.

Children's decision to explore was unrelated to the type of explanation they provided following their endorsement or rejection.

Did children's explanation for their judgment following the receipt of counter-intuitive testimony predict their decision to explore the dolls? To answer this question, we used logistic regression to regress children's decision to pick up the biggest and the smallest doll on their age (Elementary vs. Preschool) and on the type of explanation they provided. We entered the type of explanation as a set of two dummy variables with no explanation as the reference category. We did not include children who provided an explanation coded as Bigger = Heavier because only 3 children provided this explanation. Note, in Table D1, we provide the percentage of children in each age group who explored as a function of the type of explanation they provided. We display the results the aforementioned regression model in Table D2. Children's age was the only statistical significant predictor. Elementary school children explored significantly more than preschool children, controlling for the type of explanation that children provided. Explanation type was not a significant predictors of children's exploration. That is, when considered as a set, the dummy variables representing explanation type did not explain a

statistically significant amount of variation in children's search for evidence, GLH Test:

$$\chi^2(2) = 4.51, p = .11.$$

Table D1. Percentage of children in each age group who explored as a function of the type of explanation they provided.

<i>Explanation Type</i>	<i>Explored</i>
Preschool Children (<i>n</i> = 38)	
Other (<i>n</i> = 14)	36%
Bigger = Heavier (<i>n</i> = 3)	33%
Smaller = Heavier (<i>n</i> = 20)	30%
Size Sometimes Unrelated to Weight (<i>n</i> = 1)	100%
Elementary School Children (<i>n</i> = 58)	
Other (<i>n</i> = 6)	50%
Bigger = Heavier	
Smaller = Heavier (<i>n</i> = 23)	83%
Size Sometimes Unrelated to Weight (<i>n</i> = 29)	93%

Table D2. Logistic regression predicting whether children picked up the biggest and smallest dolls as a function of their age and the type of explanation they provided.

	Model 1		
	Odds-Ratios	z scores	95% CI
Elementary	5.63**	3.11	1.90, 16.70
Size Sometimes Unrelated to Weight	7.24*	2.09	1.14, 46.21
Smaller = Heavier	1.51	.69	.46, 4.95
Constant	.39	1.82	.14, 1.08
X²		30.07***	
Model df		3	
-2 Log Likelihood		88.89	

* $p < .05$; ** $p = .011$; *** $p < .001$. *Note.* $n = 93$. The explanation type “Other” is the reference category for the two variables representing explanation type. Three children who provided an explanation coded as Biggest = Heaviest were not included in this analysis.

Appendix D

Children made a judgment about the weight of the dolls: (1) immediately after the opportunity to explore the dolls by E1; (ii) when explicitly asked by E2; and (iii) when invited by E2 to select the heaviest paperweight. In table E1, we display the percentage of preschool and elementary school children who endorsed the biggest doll as the heaviest at each time point in each condition. To assess the stability of children's judgments over these three points we regressed using a multi-level logistic regression model (Stata 14's `xtlogit` command) children's judgements on the timing of these judgments using two dummy variables (Explicit Judgement with E2 and Paperweight Task with E2, the reference category was Explicit Judgement with E1). This allowed us to compare whether children's judgements changed significantly across the three time points. We conducted these analyses separately for each type of testimony.

Table E1. Percentage of preschool and elementary school children who endorsed the biggest doll as the heaviest (i) immediately after the opportunity to explore the dolls when questioned by E1; (ii) when explicitly asked by E2; and (iii) when invited by E2 to select the heaviest paperweight.

	Following Opportunity to Explore with E1	Initial Judgement with E2	Paperweight Task
Confirming			
<i>Preschool (n = 43)</i>	98%	98%	81%
<i>Elementary (n = 61)</i>	97%	100%	92%
<i>Total (n = 104)</i>	97%	99%	88%
Counter-Intuitive			
<i>Preschool (n = 38)</i>	34%	37%	45%
<i>Elementary (n = 58)</i>	48%	52%	69%
<i>Total (n = 96)</i>	43%	46%	59%

Confirming Testimony: We found that when children received confirming testimony the proportion of children who stated that the biggest doll was the heaviest did not differ whether children were asked by E1 or by E2 in a direct manner, i.e., “Which doll do you think is the heaviest?”, $z = 1.15, p = .25$. However, when asked to select a heavy paperweight by E2 children were significantly less likely to select the biggest doll relative to when they were asked in a direct manner by E1 and E2, $z = 2.54, p = .01$. However, when we tested whether this pattern applied to both preschool and elementary school children, we found that Elementary school children’s judgement that biggest = heaviest did not change significantly across the three time points. In contrast, preschool children were significantly less likely to select the biggest doll on the paperweight task relative to when they were asked in a direct manner by E1 and E2, 98% vs. 81%,

McNemar tests = .016 (see Table E1). Thus, children's judgements about the weight of the doll in the confirming testimony condition were generally stable. When asked directly about the weight of the dolls by a second experimenter they had never met before, most of the preschool and elementary school children provided answers that were similar to those they had given to E1 after they had had an opportunity to explore the dolls. Moreover, Elementary school children provided equivalent answers whether E2 asked them directly or indirectly (i.e., by asking them to select a heavy paperweight). Preschool children deviated from this pattern; they were less likely to select the largest doll as a suitable paperweight. However, the vast majority of preschoolers continued to endorse the biggest doll as the heaviest.

Counter-Intuitive Testimony: We found that when children received counter-intuitive testimony the proportion of children who stated that the biggest doll was the heaviest did not differ whether children were asked by E1 or by E2 in a direct manner, i.e., "Which doll do you think is the heaviest?", $z = 1.05, p > .25$. However, when asked to select a heavy paperweight by E2 children were significantly less likely to select the biggest doll relative to when they were asked in a direct manner by E1 and E2, $z = 5.57, p > .001$.

We followed up on this interaction by investigating whether being asked about the doll's weight differed as a function of children's age and whether they had received a prime. In Figure E1, we display the proportion of children receiving counter-intuitive testimony who stated that biggest = heaviest at three successive time-points: immediately after having had the opportunity to explore the dolls (i.e., children's final judgment with

E1); when asked by E2; and when invited by E2 to select the heaviest paperweight.

Figure E1 displays these proportions for each of the four combinations of age and prime.

We compared children's judgment that biggest = heaviest when they were asked by E1 following their opportunity to explore the dolls and by E2 (directly and indirectly) using McNemar χ^2 tests for each of the four combinations of age and prompt. We adjusted our significance level from $p = .05$ to $p = .0125$ using a Bonferroni correction to minimize the possibility of a Type 1 error. Being asked by a novel experimenter about the weight of the dolls did not increase the proportion of children who stated that biggest = heaviest (all p values $> .18$) for any of the four combinations of age and prompt. Thus, whatever claim children had made in E1's presence they also made in E2's presence.

We then evaluated whether the manner in which E2 asked children about the weight of the dolls influenced their judgments by comparing children's judgment that biggest = heaviest when E2 asked the question first directly and then indirectly (i.e., in the paperweight task) using McNemar χ^2 tests for each of the four combinations of age and prompt. We again adjusted the significance level to $p = 0.0125$. Being asked indirectly rather than directly about the weight of the dolls had a significant impact on the judgments of one group only: elementary school children who received a prompt (McNemar $\chi^2(1) = 7.00, p = 0.0082$; all other groups $p > .16$). These children went from endorsing biggest = heaviest 58% of the time to 81% of the time when asked indirectly in the paperweight task. By contrast, the remaining children made a similar judgement whether E2 asked about the weight of the dolls directly or indirectly.

Thus, children's judgements about the weight of the doll were generally stable. When asked directly about the weight of the dolls by a second experimenter they had

never met before, most of the preschool and elementary school children provided answers that were similar to those they had given to E1 after they had had an opportunity to explore the dolls. Indeed, children provided equivalent answers whether E2 asked them directly or indirectly (i.e., by asking them to select a heavy paperweight). Only one group deviated from this pattern; elementary school children who had received a prompt to explore the dolls were more likely to select the largest doll as a suitable paperweight.

Conclusion: Given the relative stability of children's judgements across the three time points in both the confirming and counter-intuitive testimony conditions, we added together the three judgements children made following the opportunity to explore the dolls (i.e., their judgement with E1, their judgement with E2, and the judgement they made as part of the paperweight task they completed with E2)

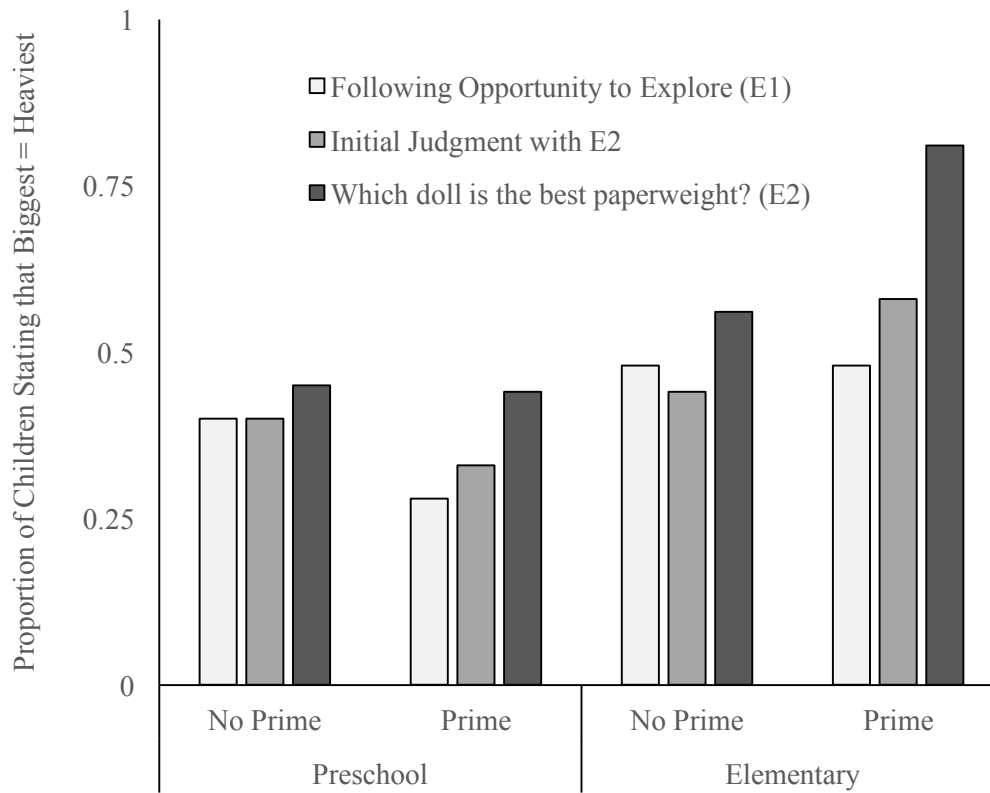


Figure E1. Proportion of preschool and elementary school children receiving counter-intuitive testimony who endorsed the biggest doll as the heaviest doll immediately following their opportunity to explore the dolls (E1), when asked by E2, and when asked to select a heavy paperweight by E2 as a function of whether children were or were not given a prime to explore the dolls.

VITA
Samuel Ronfard

2002-2006	Tufts University Somerville, Massachusetts	B.A. May, 2006
2006-2008	Pace University New York, New York	M.S.T. May, 2008
2006-2010	Middle School Teacher Department of Education New York, New York	
2010-2011	Graduate School of Education Harvard University Cambridge, Massachusetts	M.Ed. May, 2011
2011-2017	Doctoral Candidate Graduate School of Education Harvard University Cambridge, Massachusetts	
2012-2017	Teaching Fellow Graduate School of Education Faculty of Arts and Sciences Harvard University Cambridge, Massachusetts	